C.75.12
External
research

CAI UA8
-75E12

Aggregative dynamic urban models oriented towards policy

Leslie Curry
and Ross D. MacKinnon

Government Publications

FOR REFERENCE ONLY
POUR RÉFÉRENCE SEULEMENT

Ministry of State
Urban Affairs
Canada

Ministère d'État
Affaires urbaines
Canada

Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

https://archive.org/details/39210616070017

CA1 UA8
-75E12

(2)

Aggregative dynamic urban models
oriented towards policy

Leslie Curry
and Ross D. MacKinnon
Centre for Urban and Community Studies
and Department of Geography
University of Toronto

Ottawa
May 1975

FOR REFERENCE ONLY
POUR RÉFÉRENCE SEULEMENT

The views expressed are those of the authors and do not necessarily represent those of the Ministry of State for Urban Affairs. Not for quotation or reproduction without the authors' express consent in writing.

Les opinions exprimées sont la responsabilité des auteurs et ne représentent pas nécessairement celles du département d'Etat aux Affaires urbaines. Ne pas citer ou reproduire sans l'assentiment par écrit des auteurs.

Crown copyrights reserved.

Droits de la Couronne réservés.

LIBRARY
JUL 21 1977
UNIVERSITY OF TORONTO

Abstract

This study was one of the first external research contracts issued by the Research Branch of the Ministry of State for Urban Affairs. The concern then was for the methods by which social science knowledge and research could be applied to the emerging concern of the federal government with urban policy. Given the nature of the process of urbanization, the essential features of which are its complexity due to interacting factors and its dynamic nature, and the recent advancements of mathematical and quantitative approaches in social science, a particularly fruitful area of research was indicated. The present report is the product of a research contract in this area.

The report is primarily concerned, then, with quantitative models that can be used for policy development purposes. The models to meet the general requirements for this study are empirically oriented, are dynamic and, to minimize data and other problems, are aggregative. That is, they do not emphasize the decision-making processes of individual decision-making units, but emphasize the behaviour of groups of such units. The emphasis on Markov processes is a choice of dealing with dynamic models that are mathematically least intractable.

Edgar Ziegler
Director
Program Evaluation
Directorate
Ministry of State
for Urban Affairs

Abstract

This study was one of the first external research contracts issued by the Research Branch of the Ministry of State for Urban Affairs. The concern then was for the manner by which social science knowledge and research could be applied to the emerging concerns of the federal government with urban policy. Given the nature of the process of urbanization, the essential features of which are its complexity due to interacting factors and its dynamic nature, and the recent advancements of mathematical and quantitative approaches in social science, a particularly fruitful area of research was indicated. The present report is the product of a research contract in this area.

The report is primarily concerned, then, with quantitative models that can be used for policy development purposes. The models to meet the general requirements for this study are empirically oriented, are dynamic and, to minimize data and other problems, are aggregative. That is, they do not emphasize the decision-making processes of individual decision-making units, but emphasize the behaviour of groups of such units. The emphasis on Markov processes is a choice of dealing with dynamic models that are mathematically most tractable.

[illegible]
Director
Program Evaluation
Directorate
Ministry of State
for Urban Affairs

Résumé

Cette étude est le résultat d'un des premiers contrats de recherche externe passés par la Direction de la recherche du département d'Etat chargé des Affaires urbaines. Il s'agissait alors de trouver des méthodes permettant d'appliquer les connaissances et la recherche en sciences sociales, à l'intérêt nouveau du gouvernement fédéral pour la politique urbaine. Étant donné la nature du processus d'urbanisation dont les caractéristiques essentielles sont sa complexité due à des facteurs interdépendants et sa nature dynamique, et les récents progrès des approches mathématiques et quantitatives en sciences sociales, c'était là un secteur particulièrement fructueux de recherche. Le présent rapport est le produit d'un contrat de recherche dans ce secteur.

Le rapport porte donc principalement sur des modèles quantitatifs qui peuvent être utilisés pour l'élaboration de politiques. Les modèles destinés à répondre aux exigences générales de l'étude, sont orientés de façon empirique, sont dynamiques et, pour minimiser les données et tout autre problème, sont agrégatifs. C'est-à-dire qu'ils n'insistent pas sur les processus de prise de décisions d'unités décisionnelles distinctes, mais sur le comportement de groupes d'unités de ce genre. On a choisi de mettre l'accent sur les chaînes de Markov afin de traiter de modèles dynamiques qui présentent mathématiquement le moins de difficultés.

Acknowledgements

For those who know either of us it will be obvious which sections of this report were written by me and which by Les. For those who do not, we have decided to share equally the blame and the credit. The contributions of the student research assistants should, however, be made clear. We were indeed blessed with three of our very best graduate students. In the main body of the text, Russell Lee wrote Chapter II, Eric Sheppard, Chapter III and John Miron wrote the eighth section of Chapter IV. Their contributions to the Appendices are duly noted in that section of the report.

Regarding the mechanics of getting the report in its present form, we must thank Cathy Barth and Beverly Thompson of the Centre for Urban and Community Studies for typing some preliminary drafts and confronting and sometimes even successfully overcoming the obstacles of Les' handwriting.

For the financial support we must of course thank the Ministry of State for Urban Affairs. In particular Harry Swain, then of M.S.U.A., for his initial encouragement and Martin Ulrich for his patient understanding regarding our late submission. We hope that the report will be of some value to both policy-oriented researchers and research-oriented policy makers. The views expressed in this document are not necessarily those of the Ministry.

Ross D. MacKinnon

April, 1974.

Table of Contents

Preface

In a relatively short period of time, Canada has been transformed from a dominantly rural society to a nation of urban dwellers with the larger cities in particular accounting for most of the urban growth. Some have viewed this increasing importance of large cities with alarm. Many yearn for the simpler, slower paced times of the past. Others see real dangers in the concentration of political and economic power in three or four of our major urban areas. Still others perceive that large urban areas are inefficient ways of producing goods and services, diseconomies such as traffic congestion, high cost of servicing land, etc., more than outweighing the "agglomeration economies."

Many cities have undergone rapid expansion experiencing structural and functional changes in the recent past. Others have remained relatively static. Still other "urban centres" particularly at the lower end of the size scale as well as one industry towns sensitive to the vagaries of the world market for a single commodity have even decreased in size, some have virtually disappeared.[1]

The dynamic forces at work in our urban areas are numerous and varied. Perhaps the most frequently mentioned is simple growth. The number of people has increased along with the aggregate magnitude and variety of economic activity. Moreover, well known production economies of scale

---

[1] There is an ever growing literature on urban change in Canada. See for example Lithwick (1970), Bourne and MacKinnon (1972), Bourne, MacKinnon and Simmons (1973) and Bourne, MacKinnon, Siegel and Simmons (1974).

have led to the concentration of these activities, and therefore employees in these activities and their families, in the largest urban areas. These economies of scale relate to another major mechanism of urban dynamics - technological change. Changes in the nature of production activities, the proliferation of certain types of transportation, and the development of new consumer goods have sometimes generated, sometimes followed, and reinforced, new patterns of growth.

Less spectacular perhaps, but extremely important nevertheless, are the changing residential needs of families as they move through the life cycle. Much of intra-urban migration and suburbanization cannot be explained without relating it to this persistent demographic factor.

The business cycle with inflation and unemployment fluctuations is a dynamic factor very important in urban areas particularly as it relates to migration and housing starts and construction generally. Another dynamic process is the lagged sequence of investment decisions: the decision to invest in changes in the physical stock, implementation of the decision, the generally long period of payoff and deterioration, and subsequent renewal.

Indeed, many non-growth, apparently static, situations turn out to be quite dynamic just because of such time lags. Responses to a stimulus are not in general made concurrent to the stimulus. Moreover, different responses to the same stimulus occur with different time lags. These lags may be the result of poorly developed information channels or the inertia of current behavioural patterns.[2] Whatever the reason, these lags can easily generate

---

[2]Other examples of such lags are (i) the time needed for training to acquire new skills, (ii) the time required for the democratic process to operate (e.g. meeting with citizen groups, debates in Parliament, legislative assemblies, and cabinet, (iii) the time elapsed before people overcome a basic reluctance to change residences, jobs, or ways of life.

a complex, often oscilliatory, pattern of responses before an equilibrium is reached. In fact, the response of the system to out-of-date stimuli can often generate undesirable consequences. Current models of urban systems are largely static in nature. In other words, they do not involve a temporal element in any essential way. Thus if they are equilibrating models, they predict the equilibrium and not the path by which the equilibrium is attained. If an equilibrium is approached rapidly, then it is not very important whether a dynamic or static model is used. If, however, the process is attenuated over time, there is a critical difference. In fact if the system is, from time to time, the recipient of unpredictable major shocks, such as technological change, the equilibrium may never be reached. This does not invalidate the dynamic model, but may make static models completely useless. A major focus of this study then is dynamic models of urban systems - models that include time in an explicit and essential way.

Another characteristic of many urban models is their ultimate goal (often only implicitly stated) to take into account all significant factors influencing the performance of the system under study. Although this goal is never attained, the underlying premise strongly conditions the type of model which is formulated. Thus, because urban systems are evidently complex, the operational models typically assume the availability of data on many variables, and perhaps more important, a level of understanding of the nature of the disaggregate relationships between the variables that is currently unrealistic.[3]

---

[3]Whether or not one believes such an understanding is in principle unattainable is clearly dependent of the researcher's philosophy, something which is not to be discussed here. We need only claim somewhat pragmatically that this understanding is not possible now.

Errors of specification and measurement may be present for each relationship. These errors may multiply with each other rather than cancel each other out to yield a model which is neither reliable nor precise in its predictions.

The second focus of this study is to explore modelling frameworks which do not make unrealistic demands on our knowledge of the dynamics of urban systems - models which accept a large component of ignorance of our urban systems either as a current fact of life or in a more long term sense an inviolate principle of scientific research. The models are all characterized by the simple Markovian property - where the system will be in the next time period is conditioned entirely by its recent history (in the simplest case, where it is now) together with some policy decisions which are undertaken now. This is the simplest form of time dependent principle, but can yield surprisingly powerful results. In addition to being Markovian, most of the models presented are probabilistic in nature, explicitly recognizing the important random component in complex systems of all types.[4] In addition most of the models are simple in that they use only a few variables. On pragmatic grounds, it is desirable to have models for which data requirements are parsimonious. Also, it is possible to substitute a lot of information on a few variables (their behaviour over time and space) for a smaller amount of information on many variables. Where it is feasible to make this substitution, the models discussed in this study become relevant.

Finally, simple models may give qualitative insights into the behaviour

---

[4]It is interesting to note that the physical sciences, particularly physics, have adopted probabilistic modes of analysis extensively over the past 50 years.

of a system that more complex models can obscure. This study examines a number of such models and indicates how they might be augmented to be made more realistic.

It is somewhat difficult to reconcile this simplicity and randomness on the one hand with a desire to have models which are relevant to policy makers. Disaggregate (multivariate) models are apparently extremely informative to decision makers in that the sensitivity of system performance to virtually all variables (both policy and non-policy variables) can be determined. Of course, the relevance of these models may be more apparent than real if the relationships are not correctly specified or if measurement error is significant. The problem does remain, however, of how to interpret simple probabilistic models within a policy framework. Three classes of strategies can be advanced to make these models policy relevant:

First, they can be used to give predictions, limits of trends and probable deviations from these trends, given that the rules for system change do not themselves radically alter in the future. This is a fairly conventional use for many policy models. The decision maker can use the prediction in two ways: first, he can attempt to accommodate this future (population distribution, land use structure, etc.) or rather the probable ranges of values which the future may take on by building public facilities or otherwise taking public action which would seem appropriate for such futures. This basically non-interventionist, accommodating role would seem to be appropriate for many situations.

Second, one or more of the parameters of the model can be made functions of specific policy variables and the parameters of these functions fitted using conventional statistical methods. With government action (expenditures, taxation rates, subsidies, etc.) firmly embedded in the model,

sensitivity analyses can be performed to suggest which public action or combination of such actions would seem to be appropriate.

Third, an explicit goal, objective or criterion function can be appended to the model so that a measure of merit of system performance is maximized subject to certain constraints among them being the probabilistic rules of system behaviour. This third option leads us into methods of mathematical programming and optimal control theory. While these approaches are certainly useful in some contexts, they should not be viewed as attempts to eliminate the need for human decision making at all levels. They are rather attempts to increase the decision maker's ability to sort through and discard inferior policy options so a more concentrated analysis can be undertaken on a smaller number of promising candidates. Also as with all types of modelling, some fallacies of conventional, apparently common sense decisions can be illustrated.

To summarize, then, our attention is focussed on models of urban systems, characterized by a Markovian type of dynamics; in addition, we attempt to demonstrate the policy relevance of these models. Most of the models are intended to be operationalized while others give potential insights to policy without ever being made empirical.

One final comment concerning the purpose of this monograph: it is in part an interpretative review of a well established literature - a literature about which those concerned with policy implications of their research should be aware. In other sections it is on quite new, risky ground - areas in which little if any research has been undertaken in the urban context. These sections are included because work has been done in non-urban applications which shows some promise. After all, most research is imitative representing the borrowings and re-adaptations from other

disciplines and fields of interest. Certainly urban studies has been no exception to this general rule. This study in part, then represents an attempt to anticipate likely trends in future urban research.

Often the precise nature of government controls is not specified in the models discussed. Mention is made of such conventional governmental "sticks" and "carrots" as moral suasion, taxes, subsidies, absolute restrictions, and the expenditures of public money on a physical infrastructure. In most cases our level of specificity with respect to such measures is low. One reason for this is the high variability of the nature of these controls from one problem to another and even from one region or jurisdiction to another. It is probably desirable in a study such as this to keep the discussion as general as possible. Just as important, however, is our (the authors and social scientists at large) relative ignorance about the degree to which governments can and are willing to control urban development and,if they can and are willing, precisely how. The models discussed in this study assume that the willingness and capability exist. If we are sometimes vague about the precise nature of controls, this undoubtedly reflects the imprecise state of knowledge concerning controlling urban processes. The models discussed are all potentially relevant to policy related matters. If mathematical models are to give us some useful answers to dynamic urban policy questions, it is models such as these to which we will probably have to turn. One of the main objectives of this study is to initiate a discussion with policy makers to determine whether these models are relevant and how they could be made more relevant to policy questions. It is too much to ask that either policy makers or academics answer such questions independently.

Urban systems are such that our objectives for them are extremely poorly specified. The gap between the language used by the policy maker on the one hand and the mathematical modeller in the other is extremely broad and deep. This study represents one attempt to bridge this gap. The emphasis is on mathematical models which may have some relevance to policy makers. Many of the models discussed are in their present form quite simple, and adjustments would have to be made if they were to be sufficiently realistic so that major decisions could be based on them. In the models which nominally at least attempt to incorporate in an explicit way social objectives and determine the optimal sequence of decision rules, the policy makers is, quite rightly likely to be most sceptical of the models' utility. Such models can be useful in reducing the set of alternative candidates for extensive, and perhaps subjective, study, the presentation of these models is justifiable on other grounds. Mathematics can provide extremely powerful machinery which can give insights and solve problems which could not be solved using common sense methods. The ability to synthesize in an effective way mathematical models with subjective judgments is and always will be very much an art - an art which is worthy of considerable attention if the gap between policy maker and modeller is not to be continually widened.

The outsider is usually in a dilemma as to the kind of model which governments want. There is a natural tendency to argue that, being concerned with practical affairs, models should be numerical so that the "real" world is preserved. This approach is clearly valid but is also highly confining: it constricts model building to exploiting available or potentially available data and thus the type of relationship examined and the manner of its description. One might well ask whether, when governments fix policy, they require to know its results as numerical description. Certainly it is preferable, but in moving

from fixing policy on the basis of intuition and informal reasoning it is not necessary to go the whole hog in having a full numerical model. Mathematical models can occupy a middle ground for formalizing intuitive ideas and quantifying relationships without specifying them numerically. Such models have considerable advantages: not being tied to data availability, they can specify and examine relationships between concepts without the need for translating them into computational form. They can thus be used to examine fundamental notions which are important to discuss in policy making but which could not attain numerical form, save perhaps using hypothetical data. One is led to wonder whether obtaining results as a complicated set of numbers may not in fact defeat the results of the analysis. May not they hide the wood by the trees? Similarly are the numbers really necessary? For many important policy decisions, is it necessary to have a number or a probability distribution of numbers for the population of Metro Toronto in 1990 or would not a statement that it will keep on growing at least at its present rate be sufficient?

Two other issues may be raised. How far do government departments work independently? What do they regard as an infringement on their preserves? It strikes us that the really major urban policies to be determined in Canada are the same policies as those in regional development and probably those in transportation, social welfare, industries and commerce and no doubt others. Decisions are required on the need for and type of interference in the present working of the economy affecting the geography of Canada at the large scale, and its ramification in many fields. No doubt the decisions will be taken on political grounds but prior investigations via models relating various facets and examining implications would allow the decisions to achieve a viable policy within the constraints of the problem.

That a nation having a sophisticated private enterprise economy should develop and apply an urban policy implies a degree of control over the operation of some of the most basic and complex social and economic mechanisms of society. It would be advantageous if these aspects of life could be modelled within a comprehensive framework in order to ascertain how they affect each other and how they respond to various policies. Consequently the second issue to be raised is the extent to which government policies are set taking into account the full implications of that policy. In terms of modelling, how far reaching must they be? In say housing, fixing the spatial distribution of various kinds of residences in terms of existing preferences, social status, etc., is not an easy problem. However, the policy so formed will clearly affect commuting patterns: should we take urban transportation objectives into account in setting housing strategies? Ideally, of course, we should but every partial policy becomes part of a vast integrated policy, every partial model part of a general model. Even if model building could achieve this, it is doubtful if decision making could or should do so.

In discussing dynamic models having policy inputs, it is necessary to think of the sort of problems about which policies can be formed. For example, it is possible to perform sophisticated time series analysis in a multivariate framework which allows predictions to be made and controls designed. Basically these provide reductions in fluctuations through time and are thus management devices after policies have been set. Similarly we can discuss decision making but again only in the relatively short term in terms of prearranged goals. Presumably, however, the most important models are those which handle the dynamics sufficiently fundamentally that they tell us what is happening, where we are going and indicate alternatives which would be the

subject of policy discussions. It is this question of generating alternatives arising from the basic dynamics which is the hardest nut to crack. Virtually all the literature is concerned with steering towards predetermined goals with alternatives existing only in the sense of the range of projections of a narrowly defined dynamics.

In one sense, maximizing expected returns in a Markov decision process skirts many of the real problems of dealing with a stochastic process. One wonders for example what sort of trajectory the evolution of biological species would have followed in the hands of a dynamic programmer. Even if the environment be in a steady state, and if the states were denumerable, the transition rates are extremely slow relative to rates affecting genetic change. Consequently the parameters of the environment in which species would be learning and adapting to would be the intensities occurring only for relatively brief periods. The 'best' strategy may in fact be a bad strategy for the describable process, particularly the numerically describable process. On the other hand, the solution may not involve a description of an adaptive process but rather developing machinery so as to be capable of being adaptive, not in the sense of learning to be best, but in not becoming too specialized so that it is never best. This problem seems to be mathematical rather than computational or algorithmic.

There are several ways in which simple modelling can aid policy. The most obvious is system design and while there is only limited scope for this in national urban affairs (for which reason it was excluded from our terms of reference) it should not be excluded from consideration. An obvious example in Ontario is the major expansion of the post-secondary education system in recent years. It would have taken only a few simple

Markov models and some easily collected data to design a system which would meet the needs of the province much better than the present one. In design, one is not necessarily concerned with optimal policies which could well require data on cost, goals, etc. not available; simple ball-park estimates would be sufficient to try various alternatives and obtain a satisfactory solution. There must be numerous functions of this type controlled or strongly affected by governments which have a bearing on urban affairs. The system designs for each must together have an enormous effect on urban development.

The most fundamental problem bearing on policy in Canada concerns the unequal distribution of incomes due to the unequal rates of growth in various areas. The rich areas keep on getting relatively richer and the larger towns keep on getting relatively larger. This is a world-wide phenomenon with particular bearing on the political geography of national states. The basic question is whether to assume the present dynamics which are producing this result are sound. Policy would then consist of facilitating migration. Alternatively should the trend to concentration be reversed or at least reduced by one means or another? Ideally this question should be discussed neglecting political factors. The alternative should be put for $2 in Southern Ontario, say, or $1 in Nova Scotia; considerable opportunity in metropolitan Toronto or a reduced opportunity in Peterborough. Are the dynamics such that concentration can be reduced? If they are, but extra costs are involved, are these short term differences or will they be continuing? In any case, the cost and benefits should be related only to the good life for individual persons and not at all to the existing political geography. Newfoundland is a piece of earth with no greater claim than Ellsmere Island to be subsidized. The question we must ask is whether the quality of life of a child two generations from now

will be higher in Southern Ontario than in Newfoundland and how much the Southern Ontarian has to have his standard of life reduced in order to maintain the child in Newfoundland as a standard comparable to his when it could be enjoying this standard in Ontario. Is it possible to develop an urban structure in the Maritimes which would allow a reasonable population to be maintained on a self sustaining basis. Solutions to all these problems are dependent on the nature of urban dynamics. Certainly, there is not the slightest reason to believe that the vast sums spent on subsidizing development in depressed areas of the country will have the slightest long-term effects without concomitant changes in structure.

In analyses of this type, it is important to get away from bounded areas as the units with which to study space and consider rather a spatial continuum. Since towns are the basic units of development in industrialized societies, work places are obviously highly concentrated. The fact that a region lacks employment means nothing if plenty of work exists in a town half an hour or so away. Nor must we really restrict ourselves to employment within commuting distances since migration is an alternative which most of us have chosen to reach our present jobs.

These questions lead to the even more fundamental question of positivistic versus action oriented models. If its aim be to 'explain' the growth of the urban economy, a model must surely be rooted in the past and however general and timeless be the concepts it employs, it cannot but be institutionally bound to some extent. Must then an explanatory social science be tied to existing institutions when social action may require the substitution of different institutions? Put another way, must the 'scientific' content of the appropriate disciplines be largely irrelevant to policies aimed at fundamental changes in the existing structure?

To this point, the primary objectives and methodological biases of the investigators have been briefly outlined. Let us now consider the content of the report in summary form.

In the Chapter One, Markovian models are introduced so that even the reader who has very little previous knowledge of the field can grasp the basic form of the models discussed in this report. The mathematics of Markovian models are summarized and the critical assumptions identified and their implications briefly discussed. Finally, the chapter identifies in a general manner the different ways in which models of this kind can be made relevant to policy makers.

In Chapter Two, urban applications of simple Markov chain models are summarized. Applications range widely from migration to land use change, from trip making behaviour to regional development.

In Chapter Three, extensions of the simple Markov model are considered. In particular, more realistic ways of handling the temporal element are summarized. Relaxation of the assumption of stationarity and of constant intervals between transitions is discussed.

In Chapter Four, some "fundamental" approaches to the derivation of Markovian models are discussed - approaches based on basic theoretical considerations or analogies to applications in non-urban contexts. The main sources of these models are operations research, economics and physics. Queuing theory, random walks, birth and death processes, simple difference equation models and diffusion models are re-interpreted within an urban context.

Chapter Five tries to incorporate policy considerations even more explicitly in that the decision maker, his objectives, and actions are an integral part of the model. Markovian decision models are introduced together with dynamic programming solution procedures and applied to problems of

residential change. Inventory models are discussed within the context of first urban growth and change in a system of cities and then, very briefly, locational investments over time. Generalized Markov decision models are applied to an urban growth problem. Next mathematical control theory approaches are introduced, both in deterministic and statistical forms, and suggestions made for applications in a variety of urban contexts. The problem of coordinating decisions at many levels of control is identified as critical. Similarly problems of system stability are central to much of mathematical control theory. Finally, the chapter and the main body of the report closes with a maximum entropy model applied to an urban development problem.

In the preliminary work on this project a wide range of literature was studied. The long list of references in Appendix A, while perhaps not completely exhaustive, gives a good indication of the range of materials surveyed. Some of the materials have been summarized in the selected annotated bibliography in Appendix B. In the literature review, it was deemed useful to prepare papers which attempt to summarize and synthesize the major methodologies applied in specific fields. Appendix B includes of two of these surveys - one on economic urban growth models, the other on urban simulation approaches.

CHAPTER I

MARKOVIAN MODELS: AN INTRODUCTION[1]

1.1. The Simple Markov Model

In the evolution of a science, dynamic models are typically the last to be developed. Until recently there were few scientific studies of cities. Urban problems were said to be much too complex to be reduced to tractable and useful mathematical models. Certainly over the past twenty years there has been an explosion in the number of highly trained people from a number of disciplines which are attempting to do just that. There are many who would stick to the original judgment regarding the futility of such endeavors. Many if not most of the "urban modellers" themselves are quite sceptical of their results thus far achieved and some are growing less optimistic about the long run possibilities of providing definitive answers for policy makers. In spite of this scepticism on the part of both practitioner and client, there is an ever increasing number of models, some numerical, some theoretical, some both, which attempt to describe certain aspects of urban systems. Over the past few years, there has been an increasing concern with dynamic urban models, that is, with models in which time is an explicit part either because one variable is a function of time or because a variable is a lagged function of itself or other variables.

---

[1]This brief chapter is necessarily superficial. More detailed and comprehensive introductions are available in Kemeny and Snell (1960), Bhat (1972) and Howard (1971).

Perhaps the simplest and most generally applicable of dynamic models are those which have a Markovian structure. A model is said to have the Markovian property if the rules which generate the next state of the system are entirely a function of the existing and perhaps recent states of the system. It is this short memory which characterizes Markovian systems. History or at least sufficiently old history, is irrelevant in determining the immediate (and long run) future of the system. At this point, we will not discuss the substantive merits or demerits of such an assumption, only recognize that it is in fact a first approximation that considerably simplifies the analysis.

The simplest and most widely form of the Markovian property is the first order Markov assumption - the next step is influenced only by the current state of the process. The Markov terminology is usually associated with stochastic models. One of the simplest stochastic models is the finite, first order Markov chain. To clarify the exposition, let us consider a simple migration example. As we shall soon see Markov chains have been used to describe a truly wide range of urban processes, so this example should be viewed only as illustrative and not at all restrictive.

In the migration case assume that the state of the system is the geographical location of an urban resident (e.g. census tract). From time to time households change their place of residence. The first order Markov assumption states that the location of a household's next dwelling unit is entirely a function of its current location. Now such an assumption is obviously untenable except in statistical or probabilistic terms. The assumption then is making a statement about the probability distribution of the next place of residence. There is, by assumption, a family of such distributions one for every region in the system. The mechanism which drives

a finite Markov chain model is a matrix P composed of <u>transition probabilities</u> P_{ij}, the probability of moving to or occupying state or location j in the next time period given the household is currently residing at location i.

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ P_{21} & P_{22} & \cdots & P_{2m} \\ \vdots & \vdots & & \\ P_{m1} & P_{m2} & & P_{mm} \end{bmatrix}$$

The past locations of the household are assumed to give no additional information as to the likelihood of the state occupied in the next time period.

The model is finite in that there is a finite number m of non-overlapping states which exhaust all possibilities. The model is discrete in that time is assumed to be partitioned into discrete time periods; throughout any one time period the system either cannot undergo change or least the process is not observable until the next time period or stage is entered.

1.2. Markov Chain Statistics

Elementary probability theory and simple matrix algebra allow us to generate and interpret the powers of the transition probability matrix P^n. An element $P_{ij}^{(n)}$ of this matrix is the probability that a household currently in state i will be in state j n periods from now. Because of this interpretation, $P^n = \emptyset(n)$ is often called the <u>multistep transition probability matrix</u>. Because of the very short memory of the Markov process, higher powers of P yield matrices of a very particular structure. For a

process in which there is one and only one chain (a set of states from which it is impossible to escape, i.e. a monodesmic process in the terminology of Howard, 1971) we know that, $\lim_{n \to \infty} \emptyset(n)$ is equal to a matrix with identical rows composed of probabilities $[\Pi_1, \Pi_2, \ldots, \Pi_n]$. That is, for large values of n, after a long time has elapsed, it will not matter in which location a household started - a household starting in state i will have the same set of probabilities as a household starting at any other location. From a more aggregate point of view, this implies that a household selected at random will be in state j with probability Π_j - this of course implies that Π_j is the expected proportion of the population in the j^{th} zone. Thus the expected equilibrium distribution of households can be derived and is independent of the initial distribution of households. The vector Π is called the <u>steady state</u>, or <u>equilibrium</u>, <u>probability vector</u>, and can be calculated by solving a system of linear equations:

$$\Pi P = \Pi$$

$$\Pi_1 + \Pi_2 + \ldots + \Pi_m = 1.0$$

To obtain predictions of transient probabilities, it is necessary to weight the conditional multistep probabilities with the probabilities of each of the initial conditions occurring. Thus, the unconditional probabilities or proportions of population distribution after the first time period is the product of the initial probabilities or proportions with the transition probability matrix:

$$[\Pi_1(1)\ \Pi_2(1)\ \ldots\ \Pi_m(1)] = [\Pi_1(0)\ \Pi_2(0)\ \ldots\ \Pi_m] \begin{pmatrix} P_{11} & P_{12} & \ldots & P_{1m} \\ P_{21} & P_{22} & & \\ \cdot & & & \\ \cdot & & & \\ P_{m1} & P_{m2} & \ldots & P_{mm} \end{pmatrix}$$

Similarly the unconditional probabilities for the n^{th} period are:

$$\Pi(n) = \Pi(n-1)\ P$$

or equivalently:

$$\Pi(n) = \Pi(0)\ P^n$$

Of course, $\Pi(0)$ may be absolute values of population in which case the $\Pi(n)$ represents the expected population distribution after n periods have elapsed.

Because the problem has been set up in terms of probabilities, we can also easily calculate _a priori_ estimates of the variances of these predictions.

There are other summary statistics which are often encountered in Markov chain applications - statistics which may have a directly useful interpretation or alternatively are useful in a relative sense in the comparison of two or more different situations.

One of the most frequently encountered statistics of this type is the _mean first passage time_. Easy to calculate and interpret, the mean first passage time between state i and state j, μ_{ij} is simply the average number of transitions it will take the process to move from state i to state j for the first time. These statistics are easily calculated by solving systems of linear equations of the form:

$$\mu_{ij} = 1 + \sum_{k \neq j} P_{ik}\ \mu_{kj}$$

In the migration example, it could be the average number of years it will take a family to move from region i to region j for the first time. In this example, it is more a summary index of the closeness of connection between zones than an absolutely meaningful measure since any one family will not typically make moves between all pairs of zones. Note, however,

that such an index may be useful even in this case to detect important indirect connections between zones. Thus for example if one zone acts as an intermediate "way-station" for foreign immigrants the mean first passage time between the foreign state and the ultimate destination may indicate a much larger linkage than the simple transition probabilities would lead one to believe. In a system where the velocity of movement is greater, the mean first passage time statistics may have a more direct interpretation. For intracity trips, it could represent the number of stops intervening between visits from a certain land use to another (Marble, 1964); for money flows, it could be used as an estimate of the number of months it would take a dollar saved (or spent) in one city to find its way to another city, a useful index for constructing interregional multipliers (Richardson, 1973). Of course, variances of these mean first passage times can also be calculated.

It is often of interest to know how often the process will be in certain states over a specified period of time. This information can be provided by means of <u>state occupancy</u> statistics. The mean number of periods state j will be occupied over time period of length n given the initial state was state i $(M_{ij}(n))$ can be a useful statistic. These statistics are readily calculated from the multistep transition probability matrices:

$$M_{ij}(n) = \sum_{k=0}^{n} \emptyset_{ij}(k)$$

In our examples, this would tell us the expected number of time periods a household would occupy a given region, a trip maker would visit a certain land use, a quantity of money would be available for spending in a region over a specified time period given the starting state in each case. Again, variances of these statistics can also be calculated.

Another way in which simple Markov chains can differ is the rate at

which they approach the steady state distribution. One measure of this is the "shrinkage factor," (Howard, 1971). This measure is independent of the initial probability vector $\Pi(0)$ and is the amount by which the possible region of inclusion of the state probability vector is reduced in each transition. Easily calculated as the product of the absolute values of the eigenvalues of the transition probability matrix, this is a useful index of the relaxation time of the system. Of course, in practice the time of convergence towards an equilibrium will be influenced by the initial state of the system. It would be relatively easy to calculate measures of convergence for specific systems (the number of transitions to come within ε of the equilibrium vector) but no general statistic exists.

A special type of monodesmic Markov chain is one with one or more absorbing states, i.e. states which once entered cannot be left. The most obvious one, particularly in demographic models, is death. Thus even in the simple migration model, one might wish to allow for the possibility of family dissolution through death or other factors. This would certainly change the nature of the equilibrium solution - in fact without the creation of new households, all families would eventually enter the absorbing state; the equilibrium vector would consist of all zeros except for the absorbing state which would be unity. In the trip making case, the residence could be called absorbing, the trip coming to an end when one returns home. In the intercity savings flow example, investment could be the absorbing state, the process of interregional savings exchange terminating when the money is finally invested. There is a set of statistics, most notably mean times to absorption, associated with this special type of Markov chain process.

In summary then for most simple Markov chain models there is a well established, easily calculable, readily interpretable package of standardized statistics which facilitate the comparison of different Markov processes and can be helpful in providing an analyst and decision maker with a useful summary of the essential characteristics of the system in question.

1.3. Critical Assumptions of Markov Chain Models

At this point, it would be useful to enumerate the critical assumptions of simple Markov chain models. These assumptions are of course both the source of strength and weakness in these models - strength because these simplifying assumptions allow the derivation of readily interpretable, easy to calculate and quite powerful dynamic mathematical models; weakness because the assumptions and therefore the results may be at variance with the actual system under study. Virtually all models are, of course, compromises and the analyst must make the often quite subjective decision as to whether it is worthwhile to adjust a simple model so that the assumptions conform more closely to the actual situation incurring the cost of increased complexity and perhaps additional data requirements. There are five critical assumptions which have been made which are now briefly discussed and evaluated. In the Chapter III, some methods which can be used to relax some of these assumptions are presented in more detail.

The principal assumption is that the process is characterized by the Markovian property. For the first order Markov chain model, this means that only the current state of the system determines the probability distribution of the subsequent state of the system. Very few social systems have such a very short memory. For example, in the migration example, a migrant from the Maritime provinces to a central Canadian city is probably more likely to

return to the Maritimes than a lifetime resident of central Canada.

A higher order Markov chain can be formulated and solved by increasing the number of states. This has two consequences, one computational and the other relating to data requirements. If the first order transition matrix is $m \times m$, the second order matrix is $2m \times 2m$, quadruple the size necessitating more lengthy computations and larger storage requirements. Also whereas one step movements were sufficient in the first order case, now two step case histories are required, data which are often just not available.

It should be noted that even where there is some _a priori_ reason to believe a higher order Markov chain is more realistic than a first order one, the latter can often provide a very good first approximation because of the dominance of recent events.

The second initial assumption concerns the constancy of the parameters of the model. The transition probabilities P_{ij} do not change over time. In many social systems, the characteristic of _stationarity_ is unlikely to be a completely accurate characterization of the process. In the migration example, as regions change their composition, as technologies and tastes change, transition probabilities are not likely to remain constant.

Again, we have the option of using the simpler stationary model or making life considerably more difficult for ourselves by using a non-stationary model. Without going into details now (see Chapter III) we can predict transient behaviour if the probabilities $P_{ij}(t)$ are known.

The multistep transition probability is now equal to

$$\emptyset\ (n,\ t) = \prod_{i=t}^{n} (P(i))$$

Thus predictions of the distribution n stages from the starting position is

$$\Pi(0) \prod_{i=1}^{n} P(i)$$

Equilibrium behaviour in most instances is generally no longer defined and the other statistics such as mean first passage times and mean state occupancies are defined not generally but only for specified starting times and configurations. To be practical, the use of non-stationary Markov chains generally require a specification in some closed mathematical form how the P_{ij}'s vary with time. For example, migration probabilities into a region from all other regions could be positively related to the current (expected) number of people in region j. An apparent "housekeeping" problem, but an important and difficult one nevertheless is that at each stage of the process, the row sums of the matrix P (t) must be equal to unity. Tests for stationarity are also available and discussed in Chapter III.

A third critical assumption concerns the <u>discrete nature of time.</u> Changes are assumed to occur in discrete stages or at least the process can be observed only at discrete points in time. These stages need not be regular; in fact they need not be defined in terms of calendar or clock time at all. A stage in the migration example may be said to occur when a household considers or actually makes a move. In such cases it is desirable to have a model which transforms the results into real time. Semi-Markov processes can be used to achieve this result. Here the transitions as before are governed by a transition probability matrix, but the time <u>between</u> transitions is the result of a separate stochastic process which may be a function of the current state or may be a function of both current and future states.

Thus the length of residency in any area i is a random variable. Discrete time semi-Markov processes allow transitions only at regular (integral) time intervals whereas with continuous time semi-Markov processes the length of residency is a continuous variable. For numerical problems, however, the former is almost always used. Discrete time models can provide approximations to continuous time to any desired degree of accuracy, incurring of course increased computational effort in so doing.

The fourth assumption on our list is the way in which the states are defined. We have assumed that there is a <u>finite number of states which are discrete, mutually exclusive and collective exhaustive</u>. With locational states this discreteness need not be too bothersome. Although space is of course continuous locational data are usually collected on a regional or discrete basis. Where the states are measured on a continuous basis, or effectively continuous one (e.g. per capita income, city size, etc.) the finite Markov chain may introduce significant artificialities. The need to classify cities into a small number of population size groups, for example, will not only be wasteful of data, it obviously is not correct that a city at the bottom size group i has the same probability of moving to the next highest group as a city at the top of group i.

If this problem is considered serious, it may be appropriate to use more general stochastic difference equations estimating them in an auto-regressive model such as:

$$x(t) = f(x(t-1), z(t))$$

where f is a single valued function, the form of which is to be prescribed and the parameter estimated by regression methods

$x(t)$ the state of the process in time t

$z(t)$ a stochastic variable with known time dependent probability distribution.

Such models have the Markovian property and are discussed more thoroughly in Chapter IV.

The fifth and final assumption is certainly not peculiar to Markov models but should nevertheless be noted. <u>All entities in the system are assumed to be approximately homogeneous</u> insofar as they behave according to identical probabilistic rules. While difficult to determine subtle inhomogeneities, if there are strong reasons for believing certain groups in the population behave quite differently, then an adjustment to the simple model should be made. The procedure is simply to describe the behaviour of different groups with different models or perhaps with similar models with different parameter values. Thus, again to use the migration example, if households having different socio-economic status or at different stages in the life-cycle are thought to behave quite differently, then migration models could be estimated and run for each group. In this case the expected population distribution would be the sum of the expectations for each sub-model. (To make the model slightly more interesting, the transitions between the population subgroups could themselves be described according to a Markov chain.) Thus, relaxation of the final assumption is rather straightforward although detection of the inhomogeneities is sometimes difficult and the new model is considerably more cumbersome.[2]

1.4. Estimation of Transition Probabilities

The parameter estimation problem is of course a very important aspect of all modelling. With Markovian models this problem frequently raises quite substantial difficulties. Estimating the values of the parameters

---

[2]Ginsberg (1973) discusses these heterogeneity problems in considerable detail.

in the simplest model means the determination of values of the transition probabilities P_{ij}. This can be based on (a) empirical estimates from past behaviour of the system, (b) theoretical considerations, or (c) informed, but subjective judgments, even guesses.

In the simplest instance of the first case, the transition probabilities are based simply on the relative frequencies of transitions. That is,

$$P_{ij} = \frac{f_{ij}}{\sum_{k=1}^{m} f_{ik}}$$

, where f_{ij} is the number of observed transitions from state i to state j.

This is a maximum likelihood estimate of the true values of P_{ij}. It of course requires that observations are available for at least two periods -- observations on actual transitions between pairs of states for two periods or more. This is the most common estimation procedure and certainly the most preferable if suitable data are available.

Often, however, even when time series data are available observations on transitions are not. In such cases, methods which estimate transition probabilities from cross-sectional distributions or "snapshots" can be used. Optimal least squares methods of quadratic programming (Judge and Takayama, 1971) or methods which minimize absolute deviations (Scott, 1965) can be used. Theoretical considerations can be accommodated by constraining these optimizing procedures or by relating transition frequencies to exogenous variables (Spilerman, 1972b).

Also, subjective judgments regarding the values of the parameters can be made and simulating the system behaviour implied by these assumptions and subsequently making adjustments to the initial assumptions until the resulting distributions are either desirable or plausible. These sorts of simulation experiments can be used within the contexts of either the

parameter estimation problem or sensitivity analysis to proposed policy changes.

1.5. Some Additional Practical and Theoretical Considerations Relating to Markovian Models

Before considering in some detail the urban applications to which Markov models have been put, let us discuss some general attributes of this class of model - attributes which distinguish these models from other approaches to the same problems. These characteristics are not the specific assumptions which we have just outlined, but rather they are general consequences which flow from these assumptions and the way in which the models are formulated and can be used.

Markov models in general, especially Markov chain models are inherently quite flexible in terms of their format. They can be either highly aggregate models with two or three states or very disaggregate with a denumerably infinite number of states (or infinite in the autoregressive case). The level of detail is determined by two considerations: the uses to which the model is to be put and the availability of data for the model. The number of parameters increases with the square of the number of the number of states; thus data requirements (and the data must be taken from at least two time periods) may quickly limit this inherent expandible nature of Markov chain models. Of course, as we shall see not all parameters need be empirically estimated; some may be set equal to zero because of physical and legal constraints others may be derived from or restricted by underlying theoretical principles, from another model for example.

The apparent lack of causal theory underlying Markov models is frequently cited as a bothersome feature of this type of model. The degree to which Markov models can be made to relate to theoretical findings of the social sciences is a topic of much current interest in the urban literature

(Ginsberg, 1972b; Cordey-Hayes, 1972).

The theoretical basis of Markov models varies enormously from one application to another. At the very least, it represents an implicit sort of theory. Thus, some would agree that for complex systems, specific causes are difficult if not impossible to isolate. Location theory, for example, has not proven to be terribly effective as a guide to empirical research in urban systems. The theory is partial and one could argue that many statistical "explanations" are spurious. At one extreme, the use of Markov models can be an admission of considerable ignorance about urban systems as to the specific causes of change. From past system behaviour, so the argument would run, let us attempt to forecast where the system appears to be moving. This rationale makes Markov modelling a highly structured form of curve fitting. It is highly structured since the assumptions, the format, the methods of analysis and the output are very explicit. The number of ad hoc decisions is rather small.

As we shall see in Chapter IV, Markov models can be made quite theoretical in two ways. First, we can attempt to theoretically justify the parameters either by making them functions of other variables (Ginsberg, 1972b; Spilerman, 1972b) or by severely limiting the number of parameters and basing the remaining ones on quite basic and plausible assumptions. By combining two or more simple processes, one can often derive a Markov chain model. The second way in which Markov models can be made more interesting theoretically is to imbed the Markov model into a more comprehensive framework. This can take the form of simply generalizing the model to a semi-Markov process where waiting times are a function of socio-economic processes or, as Richardson (1973) does, by making the Markov model the

dynamic mechanism within a larger model which uses theoretical principles, perhaps quite deterministic, of social science.

A related criticism of Markov models is that, because of their implicit theorizing, they are not terribly useful for policy making purposes. It can easily be shown that Markov model, at least in principle, can be every bit as relevant to policy makers as the more conventional urban model. Even as a curve fitting procedure, it can be used to predict aspects of future urban systems. The means and variances together with other summary statistics are given and there is little reason to believe that these predictions are any less precise than many static, multivariate, deterministic models. These predictions can be used in the conventional way to aid in the planning of public facilities and services for the urban areas of the future.

The concern for lack of policy relevance of course lies not in this basically passive activity of accommodating future trends, but rather in how one may actively intervene in urban processes to control in some significant way future urban environments. Markov models can be useful in this context as well. The most obvious is in parametric (sensitivity) analysis. Having formulated, estimated and solved a Markov chain model, it is easy to determine the relative sensitivity of the equilibrium or some transient distribution to changes in any one or any combination of parameters P_{ij}. By experimenting with hypothesized changes in migration rates between some pairs of zones, it may be possible to identify areas in which the government has some leverage in effecting a change in population distribution at some future date. It should be noted in this context that indirect migration linkages may be quite important in influencing a desirable change. While not identifying exactly how the government agency could bring about a desired change, the Markov sensitivity analysis could aid in identifying critical migration linkages

which would yield a high payoff in moving the system towards its desired state. The sensitivity analysis of Markov model can indicate where payoffs from potential research and from policy actions will be greatest. That is, there is little point investing a lot of effort determining how P_{ij} can be changed when the forecasts are relatively insensitive to changes in P_{ij}.

Finally, in the best of all possible worlds, some or all of the parameters P_{ij} are made a function of an underlying causal structure. Indeed the Markov format could be viewed as a framework within which more conventional research efforts could be placed. At the very least it is a rational accounting model of change. Where some of the parameters are functions of exogenous variables, some of the government policy instruments such as subsidy levels, expenditures, taxation rates, etc., the Markov model has all of the advantages of the conventional multivariate, deterministic model and some of its own in addition - most notably variance estimates and a dynamic structure enabling the prediction of transient behaviour. In summary then, the Markov model allows prediction in the fact of uncertainty but also enables us to use whatever information we have concerning underlying causal structures. Additional policy relevance of Markov models is discussed within the context of optimal decision processes in Chapter V.

CHAPTER II

EXAMPLES OF EMPIRICAL URBAN AND REGIONAL MARKOV CHAIN MODELS

2.1. Introduction

The Markov chain is a relatively simple dynamic model that has had varied applications in many fields. This discussion reviews some of the empirical and empirically-oriented applications in urban and regional research. In most cases, the specific results of these studies will not be presented, however, in that they are largely unique to the particular situation. Instead, what is given is a sketch of the various purposes and strategies employed in the different studies. For purposes of discussion, the examples have been grouped according to subject matter. The arbitrary typology that was adopted is:

i) previous reviews

ii) land use change and housing stock transitions

iii) intra-urban travel behaviour

iv) functional regionalization

v) city size distribution and regional change

vi) industrial mobility

vii) social mobility

viii) migration

2.2. Previous Reviews

There have been few reviews of Markov models in urban and regional

research. Harvey (1967) has discussed the use of Markov models within a broad overview of quantitative models in human geography. And Jones (1971) has summarized some of the applications. The most comprehensive review has been by Brown (1970), although the emphasis there on functional distance, information theory and decision making differs from the tenor of this discussion.

2.3. Land Use Change and Housing Stock Transitions

These models describe the filtering and transitions in urban land use and housing stock. Predictions from the models may be used to aid in estimating future housing starts, and the need for different development and redevelopment programs, so as to attain some desired objective or "urban plan." As well, the effectiveness of various programs or policies may be evaluated by studying any subsequent responses in the urban system. In large part, however, considerations of policy implications have not been undertaken in past research. Instead, the studies have essentially been limited to description, simulation, or at best, prediction.

As one example, the approach by Bourne (1969) employed a system of multiple regression equations to compute the initial vector of the distribution of land use categories in Toronto. Some of the independent variables used included employment, socio-economic, and locational characteristics of the urban areas. The transition matrix (an average of observed conversions over two time periods) was then used for forecasting the future distribution. It was suggested that successive iterations of this method may be used for forecasting. In this instance however, estimates of the future values of all of the independent variables used in the regression procedure, must be given exogenously.

A similar framework was adopted by Stoyle (1963) to analyze the temporal

distribution of crop acreage in Pennsylvania. An observation that was made in this study appears applicable to the urban situation as well. That was that price, cost, new crop, and policy change factors could not be appreciably reflected within the stationary Markov chain, in which all of the transition parameters remain unchanged over time.

The effects of certain causal mechanisms can be postulated and tested by observing changes in the transition matrix over time. For example, Rose (1970) noted that the influx of the Negro population into business areas in Milwaukee changed the transition matrix of retail categories. What resulted was a general decline in retail activity and an increase in a few "Negro-related" activities.

Attempts to study the suburbanization process have been made by Harris (1968) for Sacremento County, California, and by Drewett (1969) for part of Berkshire, England. The Drewett study outlined a semi-Markov simulation model, in which the states were expressed in terms of percentage land devoted to urban activity in a one km. square. The simulation was achieved by two drawings of random numbers: One, to define the waiting time in a given state, described by a negative exponential distribution (so that the model reduces to a continuous time Markov-model); the other to determine the actual transition from an empirically estimated transition matrix. An obvious extension would allow for different types of urban land uses, similar to Bourne (1969); and would consider alternative waiting time distributions. Harris (1968) used a measure of accessibility, expressed as a function of population and distance to employment sites, in estimating transition probabilities from undeveloped to developed land for given urban sub-areas. In this instance, the distances to employment sites, for a given

sub-area, were considered to have a multiplicative effect. This is questionnable on two points:

i) defining the effects to be additive seems more justified; and

ii) it is unlikely that all of the different employment sites carry equal weight in determining whether a particular sub-area is to be developed; the closer centres will likely be of greater significance.

The model also assumes away the fact that the probability of development depends on how many and which other sub-areas have already been developed.

Studies more directly related to studying changes in the physical nature of the housing stock include Clark (1965), Wolfe (1967), White (1971), and Gilbert (1972).

Clark (1965) was interested in changes in average rents of rental properties by census tract for different American cities. However, the analysis was limited to visual inspection of the transition probability matrices.

Wolfe (1967) constructed transition matrices for different changes in the condition of housing over a six year period for San Francisco. Different matrices were computed for different types and ages of housing structures. It was suggested that different matrices could be constructed, corresponding to varying conditions (policies), such as a conservation program, increased code enforcement, a rehabilitation program, and so on. But at the same time, it was noted that the corresponding data requirements would be rather severe. If possible, however, such a study would offer insights into the repurcussions of alternative policies.

White (1971) was concerned with the filtering process. The act of a household leaving a housing structure was termed a "vacancy." Subsequent

moves by households into and out of this structure form a "chain," whose average length is the "multiplier." The multiplier is simply the mean time to absorption in Markov theory. It was suggested that properties of chains and the multiplier should serve as a basis for policy choices.

Gilbert (1972) has suggested the applicability of the Markov renewal process to analyzing housing turnover. The process may be characterized by the relation:

$$A_{ij}(t) = \tilde{A}_{ij} F_{ij}(t) ,$$

where $A_{ij}(t)$ the semi-Markov matrix, gives the probability of a move from state i to state j at $\leq$ time t, given a move to i at time 0.

$\tilde{A}_{ij}$ the imbedded Markov chain, the probability of an i - j move at all.

$F_{ij}(t)$ the probability of an i - j move at or before t, given an i - j move; i.e., the distribution of occupancy times in i, given the next state is to be j.

Calibration of the renewal model requires estimates of $\tilde{A}_{ij}$ and $F_{ij}(t)$, which may be obtained from longitudinal data, records of real estate transactions and so on. The model gives expressions for the expected number of units in each state, turnover rate, equilibrium vector, first passage times, number of entries into each state, and expected sojourn (length-of-stay) times in each state. Although the paper is theoretical in content, a two-state numerical example was given. The model assumes that transitions occur independently (as in the Markov chain). At the individual housing unit level, this is a rather severe assumption when one considers housing turnovers due to large development and redevelopment projects.

The renewal model is the most "advanced" of those discussed thus far. In terms of operationalization, data demands do not seem unreasonable, and analytical computation for certain types of $F_{ij}(t)$ matrices is possible. It seems, then, that a fruitful course lies in the adoption of a Markov renewal model in which the parameters are given some explicit or implicit interpretation. Of course, although this model is a generalization of the Markov chain, it still only accounts for a portion of the heterogeneity due to a non-homogenous population and environment.

2.4. Intra-Urban Travel Behaviour

Marble (1964a, 1964b) and Hemmens (1966) applied the Markov chain model to study intra-urban trip purpose linkages. Travel diaries were used to estimate the transition probabilities.

Hemmens (1964) calculated the transition matrices, and some properties derived from them, for Pittsburgh, Chicago, and Buffalo, for activity type states as well as for land use type. One purpose was to consider variations among different cities. Further applications which were suggested would consider variations due to household characteristics, results from more detailed data, time and spatial distributions related to the activities, and a generalization to a semi-Markov model.

Marble (1964a, 1964b) studies are very similar to Hemmens. An additional consideration was the comparison of the equilibrium vector to the actual distribution. This implicitly tested the existence of an equilibrium at the particular time period, which in this context implies some notion of stereotyped travel behaviour.

An interesting, and somewhat troublesome, property of the phenomenon under study is that it is cyclical (the individual always returns home), and

time-of-day dependent.

Hemmens (1966) has already remarked on the lack of an explicit time or spatial dimension in these types of studies. The latter aspect, when related to corresponding activities, is especially relevant to urban transportation planning. Another factor is the underlying decision-making process of the individual trip-maker. This may be related to work by Golledge (1967) and Golledge and Brown (1967), who visualized learning and decision-making as an information-dependent Markov process.

2.5. Functional Regionalization

The mean first passage time (MFPT) of interurban migration has been used as an index of "functional distance," which was in turn used to delimit "functional regions" nodals regions, and hierarchies. Brown and Horton (1970) used this method on selected cities in New York State.

2.6. City Size Distribution and Regional Change

Various Markov approaches have been used for analyzing city size distributions within an urban system over time.

Bourne and Maher (1970) constructed transition matrices for 5 and 10 year intervals to forecast city size class distributions in Ontario and Quebec. A similar study of Yugoslavia was completed by Fisher and Lawson (1972). In Bourne and Maher (1970) arbitrary entry figures were incremented to allow for new cities. The treatment of the incorporation and discorporation of cities, though, is arbitrary. Other approaches (Fuguitt, 1965) to this problem have added or subtracted incorporations or discorporations as they occurred, or have imputed the average of the past number of incorporations at each time interval. Of course, in a forecasting context these

methods would be limited, since they would in effect require estimates of future incorporations/discorporations.

Fano (1969) has suggested that the Markov model be used as an alternative to entropy maximizing approach to city size distributions. However, the Markov model is not really an alternative. Entropy may be regarded as a measure of randomness in a process represented by a Markov chain; and the entropy maximizing formalism may be applied to a Markov chain.

Possibly more interesting within a regional planning context is the spatial distribution of capital. Richardson (1973) has illustrated how intra-regional spatial concentration of capital (and growth) may be cumulative in the long run, through a series of Markov processes. To counter the sustained regional disparities which may be characteristic of an equilibrium situation, Smith (1961) has suggested several posibilities for increasing the incomes of some regions relative to others:

1) in a reducible chain, regions can be aggregated into irreducible submatrices; a one-time injection into the economy of one of these submatrices would then be sustained and increase income relative to other irreducible submatrices;
2) however, for one region within a set of regions in an irreducible matrix <u>continuous</u> injections are required to offset the asymptotic convergence to the equilibrium state of regional disparities;
3) alternatively, non-stationarity must be allowed, so that fiscal and monetary policies may alter propensities to trade, and hence alter the transition matrix over time.

2.7. Industrial Mobility

Labour force and industrial characteristics may serve as informative economic indicators. The application of Markov chains allows a study of changes in these indicators over time.

One such study was undertaken by Adelman (1958), in which the system states were firm size classes, as measured by value of output. And an index of industrial mobility at time n was defined:

$$I_n = \frac{\sum_{j=0}^{m} \frac{t_j}{1 - t_j}}{\sum_{j=0}^{m} \frac{S_{j,n}}{1 - p_{jj}}}$$

where t_j is the j^{th} element in the equilibrium vector

$S_{j,n}$ the j^{th} element in the distribution vector at time n

p_{jj} the jj^{th} element in the transition matrix.

However, a problem when using value of output as a measure is that it must be adjusted over time to allow for inflation and industrial advances.

Instead of considering the changes within firms, Denton (1973) has looked at month-to-month labour force movement in Canada. The state variables in the transition matrices (for different sexes) denoted employment, not in labour force, and various states of unemployment. The parameters were econometrically estimated from equations of the form,

$$P_{ij,t} = a_{ijo} + \sum_{k=1}^{11} a_{ijk} D_k + a_{ij\,12} \overline{U}_t + a_{ij\,13} \Delta U_t$$

where D_k is a dummy variable representing the months of the year except December (which was not included so as to avoid the

problem of matrix singularity)

$\overline{U}_t$ a moving average, $\frac{1}{2}(U_t + U_{t+1})$

ΔU_t first-order difference, $U_{t+1} - U_t$

U_{t+1}, U_t unemployment rate at times (t+1) and t (these were not used in the regression estimation due to the high correlation between them.)

The model has been extended by disaggregating by age, and by region - although these results were not reported. In addition, it would be informative to disaggregate the "employment" category into specific occupational categories, as in Blumen, et al. (1955). The main shortcoming of the Denton (1973) model appears to be the fact that its use is restricted to simulation rather than prediction. Prespecified values of U_{t+1} are required since these are not estimated endogenous to the model.

Blumen, Kogan and McCarthy (1955) studied the flow of manpower between occupational categories. The main importance of this study lies in its innovative treatment of one aspect of population heterogeneity via the "mover-stayer" model. The population was partitioned into "stayers," (who were regarded as not changing their jobs at all) and "movers" (who were all considered to behave according to the same decision process, represented by a Poisson probability distribution. This revised model produced a better fit when compared to the actual transition matrix than the usual Markov model, which underestimated the main diagonal elements. This latter property forms the basis of the "law of cumulative inertia" (McGinnis, 1968) and subsequent modifications of the Markov model in migration and social mobility models.

2.8. Social Mobility

Conceptually, Markov models of social mobility are extremely similar to those used in migration and industrial mobility studies. Differences stem largely from variations in the definition of the states. For example, the study by Adelman (1958) that was previously discussed, followed the approach in Prais (1955). In the latter, a transition matrix was computed from data from a random sample of father-son pairs, with the parameters representing the probability of the son being in the j^{th} occupational class, given that the father was in the i^{th}.

Of course, it is easily realized that the Markov model as applied in most of the studies discussed thus far is somewhat restrictive and unrealistic. To allow for population heterogeneity and changes over time (non-stationarity), several theoretical extensions have been suggested within a social mobility context. McFarland (1970) has suggested that time-stationary Markov matrices be specified for <u>each</u> individual. However, this would encounter some rather severe problems when one considers the astronomical parameter estimation demands.

At a more aggregated level, Stone (1972) has suggested that transition probabilities might be allowed to vary with the past transition history. This idea appears to reduce to a case of the cohort-survival class of models that have been outlined in some detail by Rogers (1971). Stone (1972) also suggested another method for treating non-stationarity, which imputs an iterative matrix operator on the transition matrix. However, the operator was itself assumed to be stationary, so that changes in behaviour are themselves assumed to follow an unchanging trend.

2.9. Migration

The somewhat arbitrary distinction has been made in this discussion between industrial mobility, social mobility, and migration studies. The first type may be identified by the fact that the definitions of the states relate to industrial or labour force characteristics; the second, socio-economic or demographic attributes; the third, some explicit locational specifications and hence some notion of spatial interaction. General discussins relating the Markov model to spatial interaction include those by Olsson and Gale (1968) and Gale (1972). Detailed Markov-related matrix accounting models have been presented in Rogers (1968, 1971).

Some relatively straightforward applications of the Markov model include those by Tarver and Gurley (1965), in which the transition matrix computed from 1955 and 1960 data was used to project estimates of population distribution in 1955 and 1970 for the nine census divisions in the United States; and by Lindsay and Barr (1972), who used a gravity expression on data from a questionnaire survey sample to estimate the Markov parameters. In the latter study, however, a different transition matrix was computed for each time period, so that the model in effect loses its forecasting power.

The remaining models which are discussed are extensions of the simple Markov chain. Although several of the papers are theoretical it is felt that they are representative of the types of models that are likely to be considered whenever the simple Markov chain model proves inadequate.

One thrust in mathematical sociology has been aimed at generalizing the Blumen, et al. (1955) mover-stayer model. One such paper by McGinnis (1968), presented an exposition of the "Cornell Model," whose

central property is the "Axiom of Cumulative Inertia." This postulate states that "the probability of remaining in any state of nature increases as a strict monotonic function of duration of prior residence in that state" (McGinnis, 1968, p. 716). This notion was portrayed by including an additional dimension (predictor variable) in the transition matrix. However, the model reduces to Markovian transitions within duration-specific classes. Henry, et al. (1971) have modified the McGinnis (1968) model by placing a finite upper bound on the duration-of-stay effect, so that the probability of leaving a location becomes constant after a certain duration time. This guarantees a finite state space, hence allowing solutions. However such a scheme is somewhat difficult to rationalize in the context of the actual process. Morrison (1967) and Land (1969) have attempted to verify the duration-of-stay postulate using Amsterdam, The Netherlands and Monterrey, Mexico data. However, the results of their regression analyses were not convincing. For some of the parameters, there was a wide difference in the parameter estimates (even different signs) between Amsterdam and Monterrey (Land, 1969). It appears reasonable to suggest that duration-of-stay effects are likely to differ significantly between different locations and especially for different phenomena (e.g. migration vs. housing turnover).

Spilerman (1972a) has engaged the problem of population heterogeneity by using one individual-level transition for the whole population, but allowing individual mobility rates to vary. Hence, if individuals' transitions follow a Poisson process and if individual rates of mobility follow a gamma density, then a negative binomial will describe the population's movement frequency-distribution. In an empirical example on interregional migration in the United States, several operational problems emerged:

i) in the analysis, people with greater than four moves were omitted, eliminating the extremely mobile portion of the population;

ii) there was little interregional migration since only four states were used; such aggregation limits the information gained from the analysis;

iii) respondents underestimated the number of moves when asked about the distant past; and

iv) the assumption of stationarity in the transition matrix was violated, largely due to life cycle changes.

An important approach that has already been noted (see the discussion on Gilbert's (1972) model) is the semi-Markov or Markov renewal model. Ginsberg (1971) has discussed its appropriateness for migration, and has shown that the "cumulative inertia" axiom of McGinnis (1968) may be accommodated within the semi-Markov framework. Ginsberg (1972a) has also stressed that unless parameters of probabilistic models are related to the causal structure and exogenous determinants of the (migration) process, such models will have little substantive or theoretical content. He then suggested (Ginsberg, 1972b) how the transition probabilities could be estimated from a gravity model, which he shows to be compatible with Luce's (1959) choice theory.

Other models that give transition probabilities or transition rates as a function of the sizes of the origins and destinations have been presented by deCani (1961) and Cohen (1972).

What Ginsberg's (1972a) suggestion implies is the introduction of independent variables into a Markov model. Denton (1973) and Bourne (1969) are two of the models already discussed that adopt this strategy. Spilerman (1972b) suggests using past residential location and socio-economic attributes of the population in estimating the transition matrix; whereas Kelley and

Weiss (1969) and Morrison (1973) present econometric estimation formats using wage differences, degree of unemployment, etc., as independent variables.

2.10. Concluding Remarks

In conclusion, several points can be noted:

a) For a given Markov process, characterized by its transition matrix and initial distribution vector, its associated properties (predictions by powering the matrix, mean absorption time, mean first passage time, equilibrium vector, etc.) may be computed. This possibility is implicit, even though most Markov chain applications have concentrated on selected aspects of these properties.

b) In most instances, tests of the appropriateness of the Markov property to the phenomenon under consideration have been done only *a posteriori*, by comparing the actual state to the theoretical prediction, rather than *a priori*.

c) The semi-Markov or Markov renewal model provides a general treatment of duration-of-stay effects, and is theoretically preferable to the Markov chain. At the same time, it should be realized that the semi-Markov model accommodates only a small fraction of the various sources of population heterogeneity.

d) It has been stressed that some theoretical, i.e. causal, interpretation be attached to the parameters. It was also noted, however, that some of the regression-type approaches to this problem required exogenous specification of future values of some of the independent variables. Hence the solution to one estimation problem poses another.

c) In terms of planning, it would be desirable to have a model in which the parameters were made sensitive to "controls" or "policies." These in turn would be determined according to some specified objective function or urban/regional plan.

All of these issues are explored in following sections of this report.

CHAPTER III

RELAXING SOME OF THE ASSUMPTIONS OF SIMPLE MARKOV MODELS

3.1. The Treatment of the Temporal Element in Markov Models

Given that the present situation of a process can be specified by some measured value, and given the set of all possible such values; then a Markov process can describe the movement of the system from one such value, or state, to another in successive time periods. This movement is determined by specified probabilities of travelling between each pair of states, known as transition probabilities, which are defined for all time periods and all states. It can be seen that the model is dynamic in the sense that time is an explicit element, and indeed the model is such that the expected state of the process can be described at any time, given the initial state and the transition probabilities. So the time-path of the process can be followed in detail.

As with all mathematical models, however, the model is not tied to any particular empirical theory, and its application always depends on the validity of the mathematical process as an analogue of the observed one. This in turn depends on the specific assumptions of the model; which are explicit because of its mathematical nature, but which are not necessarily a reasonable representation of the real world simply because of this.

In this chapter, it is proposed to examine the assumptions in Markov models which define the representation of time and its effects on the evolution of the process. There are three major assumptions involving time which are made in the simple model.

a) Time is seen to be discrete; with the system undergoing a change in state (even if this involves moving from a given state back to itself

again) at the end of every unit of time.

b) The process is assumed to be Markovian; meaning that the present state of the system depends on the state it was in immediately previously and only on that state. This "no memory" assumption has undergone much criticism.

c) The process being modelled is assumed to be stationary, implying that conditions specifying the behaviour of the system at any one time period are the same as those for all other time periods. This is the thorniest of these assumptions to be overcome.

The Markov model contains other assumptions unrelated to time, but these will not be discussed here.

This chapter has four parts. First, methods of testing statistically for the existence, in empirical data, of conditions satisfying the second and third assumptions will be presented. (The first assumption is not really testable, as in practice time is not discrete and processes rarely act as if it were. The validity of this assumption can only be estimated when the model is applied to a particular data set.) Second, attempts to incorporate violations of the temporal assumptions within the simple Markov framework will be discussed. Then, the means available for modifying the first two assumptions within the semi-Markov model will be covered. Finally an account will be made of attempts to model non-stationary Markov processes. Within these mathematical modifications, there is also the question of changing data needs in order to operationalize the models; and discussion of the questions of the trade-off of extra work and data for a gain in conceptualization of the process will be included.

3.2. Statistical Inference and Markov Processes

It is often required to make tests on observed data to see if the process it represents obeys the assumptions of Markov models. The data usually comes as a set of frequencies of occurrence of the system in different states and time periods. Given k states and T time periods, then T matrices can be constructed of the form:

$$
\begin{array}{c|ccccc}
\text{Origin States} \setminus \text{Destination States} & 1 & 2 & . & . & k \\
\hline
1 & f_{(11)}(t) & f_{(12)}(t) & & & f_{(1k)}(t) \\
2 & f_{(21)}(t) & f_{(22)}(t) & & & f_{(2k)}(t) \\
. & & & & & \\
. & & & & & \\
k & f_{(k1)}(t) & f_{(k2)}(t) & & & f_{(kk)}(t)
\end{array}
$$

Where $f_{ij}(t)$ is the observed frequency of occurrence of movement from i to j between time periods t and t+1.

The theoretical basis for hypothesis construction is that the distribution of the $f_{ij}(t)$ tends to be according to the multinominal distribution. Then the maximum likelihood estimates of the transition probabilities from any i to any j in the time period t to t+1, $p_{ij}(t)$ are:

$$p_{ij}(t) = \frac{f_{ij}(t)}{\sum_{i=1}^{k} f_{ij}(t)} \qquad \text{(Anderson \& Goodman, 1957)}$$

On the basis of this, tests can be constructed to measure deviations of observed frequencies from structures expected under the null hypothesis, and to say when such deviations become statistically significant. Two types of test have been developed; the χ^2 contingency table test, and the maximum likelihood ratio test. The two are asymptotically equivalent,

but have different small-sample properties. There is little agreement on which of the two is more powerful, but Kullback et al. (1962) state that the latter statistic has the property of being additive. If two hypotheses can be added to give a composite hypothesis, then a composite test for this is simply the addition of the tests for the two original hypotheses. In addition calculation is easy.

It is this latter statistic which will be used here. It is formally known as the minimum information discrimination statistic, and is equivalent to $-2\ln L$, where L is the maximum likelihood ratio. The general form of the statistic is:

$$2I = 2 \sum_{i=1}^{k} f_i \ln \frac{f_i}{np_i}$$

Where there are k categories and,

f_i = observed frequency in ith category.

p_i = prob. of ith category under null hypothesis.

n = total sum of all f_i.

3.2.1. Testing for the Existence of Markovity

This falls into two parts:

i) Zero v. first Order.

This is to test whether the sequence of events is better described by a process where future occurrences are independent of past states, or by a first order Markov process. The null hypothesis is that for each time period, the probability of being in any state j is constant, against the alternative hypothesis that this does not hold.

i.e. H_o: $p_{1j} = p_{2j} = \dots = p_{kj} = p_j$ $\quad$ p_j = prob. of being in j

or that $P_{(j|i)} = P$ $\quad$ for all i

i.e. $\dfrac{p_{ij}}{p_i} = p_j$

or $p_{ij} = p_i\, p_j$

Kullback et al. (1962) give the test for this as:

$$2I = 2 \sum_{i=1}^{k} \sum_{j=1}^{k} f_{ij} \ln \frac{f_{ij}}{\frac{f_i f_j}{n}} \qquad f_i = \sum_j f_{ij} \qquad f_j = \sum_i f_{ij}$$

$$= \sum_i \sum_j 2f_{ij} \ln f_{ij} + 2n \ln n - \sum_i 2f_i \ln f_i - \sum_j 2f_j \ln f_j$$

The test is asymptotically distributed as χ^2 so the result is determined from χ^2 tables with (k-1).(k-1) degrees of freedom.

ii) Second vs. First Order Markovity.

If a traditional Markov process model is going to be used, it is necessary that the data not only show dependence on a previous state, but also that this is only first order dependence. To test that this is true, data is necessary for two time periods, so that frequencies of movement between three successive states can be calculated.

Here the null hypothesis is that the probability of moving from i to j to k is the same as the probability of moving from j to k on the second transition, no matter which state was occupied previously.

i.e. $P(k/j,i) = P(k/j)$

The alternative hypothesis is that the probabilities vary depending on the state occupied two time periods ago.

The test here is:

$$2I = 2 \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} f_{ij\ell} \ln \frac{f_{ij\ell}}{\frac{f_{ij} f_{j\ell}}{f_j}} \qquad f_{ij} = \sum_{l} f_{ij\ell}$$

$$f_{jl} = \sum_{i} f_{ij\ell}$$

$$f_j = \sum_{i} \sum_{l} f_{ij\ell}$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} 2f_{ij\ell} \ln f_{ij\ell} + \sum_{j} 2f_j \ln f_j - \sum_{i} \sum_{j} 2f_{ij} \ln f_{ij} - \sum_{j} \sum_{l} 2f_{j\ell} \ln f_{j\ell}$$

Where $f_{ij\ell}$ is the number moving $i \to j \to \ell$.

The test is looked up at $k(k-1)^2$ d.f. It is easily calculated given tables and the observed frequencies.

For the first order condition to hold, the null hypothesis in part (i) must be rejected and in part (ii) accepted.

3.2.2. Testing for Homogeneity

Here, under the null hypothesis the individual probability matrices constructed from the frequency tableaus for each time period would all be equal. Under the alternative hypothesis at least one of the matrices would differ from the others.

The test for this is equivalent in form to the test for first v. second order Markovity (Kullback et al., 1962). The test is:

$$2I = 2 \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} f_{ij}(\ell) \ln \frac{f_{ij}(\ell)}{\frac{f_j(\ell)\, f_{ij}(.)}{f_j(.)}} \qquad f_j(1) = \sum_{i} f_{ij}(\ell)$$

$$f_{ij}(.) = \sum_{l} f_{ij}(\ell)$$

$$f_j(.) = \sum_{i} \sum_{l} f_{ij}(\ell)$$

Which has $k(T-1).(k-1)$ d.f.

Acceptance of the null hypothesis would indicate the existence of homogeneity.

Only two separate time periods of data are necessary to carry out this test; but obviously the more periods available, the greater the confidence that could be put in the results.

When the assumptions are deemed to be satisfied, then the discrete-time homogeneous Markov process may be a valid representation of the observed phenomenon.

3.3. Modifications within the Simple Markov Model

Before describing the modifications that have been made within the traditional format, it will be useful to write down the mathematical structure of the simple model as described in the Chapman-Kolmogorov equations.

Given $Pr(s(t)=j/s(t-1)=i)=P_{ij}$ for all t (as the process is homogeneous) then for a process containing n states,

$$Pr(s(t)=j/s(t-2)=i)= \sum_{k=1}^{n} P_{ik}.P_{kj}$$

Now let $q_{ij}(t,T)$ be the probability of travelling from i to j in the time interval t to T: the multi-step transition probability

then

$$q_{ij}(t-2,t) = \sum_{k=1}^{n} P_{ik}.P_{kj}$$

and in general,

$$q_{ij}(t,T) = \sum_{k=1}^{n} q_{ik}(t,r).q_{kj}(r,T)$$

for any r,

or if we define $Q(t,T)$ as the matrix of $\{q_{ij}(t,T)\}$, and P as the matrix of $\{P_{ij}\}$; then:

$$Q(t,T) = Q(t,r).Q(r,T)$$

Now if we let $r=T-1$;

$$Q(t,T) = Q(t,T-1).Q(T-1,T)$$
$$= Q(t,T-1).P$$

This is the Forward Chapman-Kolmogorov (C-K) Equation.

Similarly we may derive the Backward C-K Equation by defining r as t+1 :

$$Q(t,T) = P.Q(t+1,T)$$

Because these are homogeneous processes, only the differences T-r and r-t matter and not the actual values of T, t, and r. If, say, y=m-t=T-r then Q(t,m) would be identical to Q(r,t), and both could be defined as Q(y). Then if t=0, the Forward C-K Equation could be written as,

$$Q(T)=Q(T-1).P$$

with solution:

$$Q(T)=Q(T-2).P.P = P_{(o)}\ P^{T}$$ Where $P_{(o)}$ is the initial distribution.

3.3.1. Dealing with Continuous Time

To convert to a continuous-time model, all that is needed is to restate the C-K equations in continuous form (Bharucha-Reid, 1960).

For this section we will redefine P_{ij} as the probability of going from i to j in time t+dt; and $u_i(t)$ as the probability of a change occurring in t+dt, given the process was in i at time t. Then the differential equations may be derived (Bharucha-Reid, 1960);

The forward equation,

$$\frac{\partial q_{ij}(t,T)}{\partial T} = -u_j(T).q_{ij}(t,T) + \sum_{l=1}^{k} u_l(T).P_{lj}.q_{il}(t,T)$$

and the backward equation,

$$\frac{\partial q_{ij}(t,T)}{\partial t} = u_i(t).q_{ij}(t,T) - \sum_{\ell=1}^{k} u_i(t).P_{i1}.q_{1j}(t,T)$$

Now if we let B be a matrix where $b_{ii} = u_i$

and $b_{ij} = u_i P_{ij}$

assuming $u(t)$ constant for all t.

Then we have a matrix of infinitesimal transition probabilities, equivalent to P in the discrete case.

Then the forward equation becomes

$$\frac{\partial Q(t,T)}{\partial T} = Q(t,T).B$$

and the backward equation;

$$\frac{\partial Q(t,T)}{\partial t} = B \cdot Q(t,T)$$

The solution to the latter is:

$$Q(t,T) = P(t).e^{A(T-t)}$$

where $P(t)$ is the state of the system at time t.

This is equivalent to the discrete case;

$$Q(t,T) = P(t) \cdot P^{(T-t)}$$

Although stated in continuous form, the Markov model has not lost any of its powers of solution; and like the discrete form expressions may be derived for properties such as limiting vectors and mean first passage times. However there is little gain conceptually in the model in that probability of change is still constant for all time periods, giving rise to similar expected solutions to those of the discrete case. The only new element is that change does not have to occur at the end of finite units of time. Here the probability of moving from a state is an exponential function of time

(Ginsberg, 1971), $ce^{-c.t}$ where c is a constant that depends on the state. This is related to the solution to the differential equations given above. However, the major conceptual problems in fitting the Markov model to current theory lie with the other assumptions.

3.3.2. Duration of Time in a State and the Mover-Stayer Problem

The major discussion of the utility of Markov models of social behaviour has centred round problems of modelling geographic and social mobility. Here one of the principal difficulties is the duration of residence effect, by which probability of movement is seen to be related to the length of time spent in the present state. However, the simple Markov model assumes probability of movement is only related to the previous state occupied (the Markovity assumption).

Blumen, Kogan, and McCarthy (1955) demonstrate that the simple model underestimates the number of people staying in any given industrial category after a certain period of time. To help account for this cumulative inertia effect, they split their population into two groups; the stayers who never move; and the movers who are all subject to the same transition probability matrix. This improved the fit of the model to data.

An alternative approach to this same problem was to keep the population homogeneous (i.e. all subject to the same probabilities of movement), but to define a different probability matrix for each allowable sojourn time in a state. Thus the population was divided up in terms of its past history (expressed as the length of time since an individual last moved), rather than in terms of different inherent mobility characteristics of subgroups of this population. This approach, taken by McGinnis (1968), effectively redefined the concept of state to keep the model within the traditional matrix form. Given k locations between which people can move,

and T different duration times (0, 1, 2, . . ., T), then for each location every duration time represents a different state. Thus if a person in location 1 who has been there for t timeperiods is still there one period later, he has effectively moved from state (1, t) to state (1, t+1) which means he is subject to the t+1st probability matrix now rather than the tth one. Here, a different matrix for each duration period is needed which implies large data requirements unless the probability values are linked to theory. In addition this model is difficult to solve, and the authors have had to resort to inconclusive simulation experiments (Ginsberg, 1971). The model is still homogeneous in the sense that none of the matrices vary through time.

Henry et al. (1971) modify the model by putting in a maximum duration time, after which either people do not move (giving rise to an absorbing process), or else they move according to a matrix which is constant for all duration times greater than the 'maximum' (giving rise to a regular Markov process). To further aid analysis of the process, the model is split into two parts; the probability of moving and the probability distribution of where the move will go. This approach will be illustrated later in this chapter.

In response to difficulties with this type of model, McFarland (1970) points out:

> . . . the observed decline in mobility rates over time does not require an explanation involving a corresponding reduction in mobility probabilities . . . [it] can also be explained by heterogeneity of a population in a model where each person's transition probabilities are constant over time.

In this case each person operates according to his own simple Markov model. This is really the model of Blumen et al. carried to its logical conclusion.

One result of this was a new expression for the multi-step transition probabilities.

Defining:

$N_o(m)$ as a diagonal matrix with a one in the position of the mth individual's original location, and zeros elsewhere.

$$N_o = \sum_m N_o(m)$$

$P_{(m)}$ as the transition matrix of the mth individual.

Then:

$$P = N_o^{-1} \sum_m N_o(m) \ P_{(m)}$$

is the one step probability matrix,

and,

$$Q(o,t) = N_o^{-1} \sum_m N_o(m) \ (P_{(m)})^t$$

is the multi-step probability matrix.

For the simple model, the multi-step matrix $Q(0,t) = P^t$ assumes that the probability of an individual remaining in the same state for two time periods is the square of the probability of being there for one time period. This has typically underestimated the numbers staying on for long periods (Spilerman, 1972). The above model will help correct for this. The McFarland model will produce the cumulative inertia effect without explicitly including it, because those who leave a state first tend to be the more mobile, meaning those that are left are more likely to stay even longer. In this way, there is a positive correlation between duration of stay and disinclination to move, simply due to the inherent immobility of certain parts of the population.

Spilerman (1972b), in an important paper, has advanced from here; attempting to find independent variables that will help predict transition probabilities and thereby cut down on data needs. Given a composite

transition matrix for a population, and given individual and social characteristics of the individuals moving between the states, a multiple regression model is constructed to give a best fit prediction of probability of movement for any values of the population characteristics. One predictor equation is set up for each pair of states and all the equations are then put together to form a general predictor. From this, equations can be taken out for sub-populations which possess just certain attributes; each with a predicted transition matrix inferred from the multiple regression equations. This provides a methodology for disaggregating the population into a set of heterogeneous groups.

Two approaches have been tried in an attempt to modify the Markovity assumption and take account of the effect of duration time on movement. Both require greatly increased data needs, but of the two the population heterogeneity approach seems to be the more practical; especially as it seems easier to predict transition probabilities from information on human attributes than to try to define a duration time effect that will hold for the entire population. To predict this latter effect requires data directly relating the two variables, with others held constant, before it can be tested, and in addition this effect may vary with different types of mobility (Spilerman, 1972b).

3.3.3. Allowing for Inhomogeneity

Spilerman (1972b), in a further application of his theory linked approach to Markov probabilities, distinguishes the effects of temporal inhomogeneity that are a result of structural change (due to changes in parameters that specify model relations), from those effects that are due to natural demographic change in the population (since age can affect mobility). Under the latter type of change, the regression constants would still be valid and could be used to predict new migration probabilities

given demographic change; whereas under structural change these parameters would have to be recalculated. Spilerman gives an example comparing the effect of using regression coefficients and the given demographic change to predict transition probabilities and the movement of population; versus the effect of using the observed one-step transition probabilities from each time period multiplied together to give a prediction. Of the two approaches, Spilerman's model works better. Also, he shows that by comparing the results of such a model with the observed migration, an idea can be obtained of how highly structural change is affecting the process as opposed to demographic change. Ideally if it were all due to demographic change, the modelled and observed results would be very similar; whereas structural change, which is not allowed for, would lead to greater discrepancies. This is the first time that it has been attempted to link substantive empirical theories to the abstract mathematical Markov model; a vital problem in the further empirical use of such models (Ginsberg, 1972a).

The only other attempt to incorporate inhomogeneity into the simple Markov model has been by Ginsberg (1971), who essentially redefined the use of the concept of time. He describes 'operational time,' which he defines in the following form:

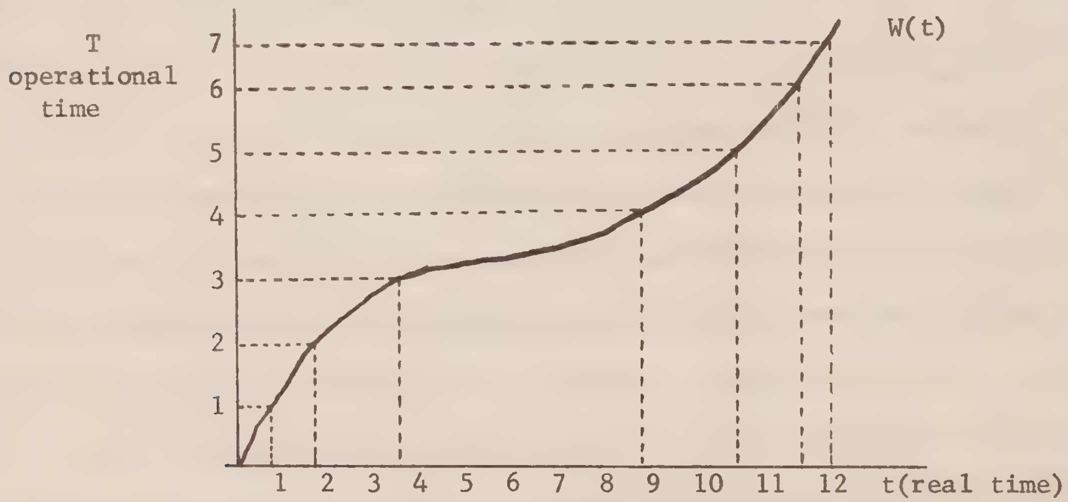


Figure 3.1

Given a function $T=W(t)$, where $W(t)$ is, say, the number of individuals, T, moving into any state j in length of time t, then this function might vary in a non-linear form with t as shown in Figure 3.1. This might be because mobility varies with age which itself is a linear function of real time.

Now T can be defined as a new time scale such that when $W(t)$ first reaches one, $T=1$; when $W(t)$ first reaches 2, $T=2$ and so on. Thus in this operational time scale the mean number of events in a time interval of length u is always u by definition, implying that the probability of entering a state is homogeneous with respect to this time scale. Relating this to real time, if $W(t)$ were increasing at a decreasing rate with respect to real time, this would imply that the probability of the event occurring will decrease over time. If $W(t)$ is accelerating its rate of increase, then the reverse is happening. Thus in the example above, as an individual grows older his probability of entering j at first decreases and then increases -- an inhomogeneity which is smoothed out if the model is run in operational time.

The use of this procedure needs a theoretical rationale for relating some homogeneous time scale in which the model will operate to real time, and also requires data to calibrate the relationship. But if this is possible, then the model can be run in operational time and its results readily translated back into real time. However when inhomogeneity is due to the influences of more than one variable, this approach becomes difficult.

Although ways of getting around inhomogeneity are outlined above, they give no idea of the properties of the model when inhomogeneity is introduced - particularly structural inhomogeneity. If a process is truly inhomogeneous, then for the Markov model to be of some use it must be possible to get some idea of how the process changes in real time.

3.4. Relaxing Markovity: Semi-Markov Processes

After reviewing the current state of Markov models in mobility research, Ginsberg states:

> Rather than trying to force the analysis into the framework of Markov chains, I reformulate McGinnis' assumptions so that the mobility process he envisages becomes an instance of what is known as a Semi-Markov or Markov Renewal Process. (Ginsberg, 1971, p236).

In the semi-Markov model, the Markovity assumption is modified so that the probability of change depends on the duration of stay in that state, and also on the states that the process is moving from and to. So dependence on the future as well as the past is built into the model.

For this purpose, a new random variable is formulated; the holding function. This is defined as:

$h_{ij}(m)$ the probability that the process remains in state i for m time periods before moving to state j.

The choice of state to move to is still determined by a homogeneous transition probability matrix P. In effect the model first uses P to select a destination state and then selects the appropriate holding function to determine when the move is likely to take place.

In addition to the holding function, a waiting function can also be defined as:

$W_i(m)$ the average probability of remaining in state i after m time periods have elapsed, no matter what the destination.

This is equal to:

$$W_i(m) = \sum_j P_{ij} h_{ij} (m)$$

Time can be continuous or discrete, it makes little difference to the computation of the model. The holding functions may be of any mathematical form. We may also define probabilities of movement by the time m time periods have elapsed:

$$h'_{ij}(m) = 1 - h_{ij}(m)$$

and

$$W'_{i}(m) = 1 - W_{i}(m)$$

Now the overall probability of moving from i to j by the time the model has been in state i for m time periods, $a_{ij}(m)$, may be specified:

$$a_{ij}(m) = h_{ij}(m)\, P_{ij}$$

Here the probability matrix specifying P_{ij} is known as the <u>embedded</u> or <u>corresponding Markov chain</u>.

Multi-step transition probabilities can now be derived (Howard, 1971):

$$q_{ij}(t,T) = \delta_{ij}(W_i(T-t) + \sum_{r=t+1}^{T-1} \sum_{\ell=1}^{k} P_{i1}\, h'_{i\ell}(r-t).q_{\ell j}(T-r,T)$$

$$t < r < T$$

$$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}$$

$$= \delta_{ij}(W_i(T-t)) + \sum_{r} \sum_{\ell} a_{i\ell}(r-t).q_{\ell j}(T-r,T)$$

Where the first part of the expression gives the probability of the process not leaving state i in the time interval t to T. In this model a convolution summation has to be taken over all intermediate times r, as the probabilities are no longer independent of the length of time interval r-t spent in a state.

In matrix form, the above equation becomes:

$$Q(t, T) = W(T-t) + \sum_{r=t+1}^{T-1} A(r-t).\, Q(r, T)$$

Where W(T-t) is a diagonal matrix specifying probabilities of remaining in any state i for r-t time periods, and A(r-t) is the matrix of probabilities of movement between states in r-t time periods.

Replacing summation-type convolutions with integration over all r would effectively convert the process to continuous time.

This type of model contains many similarities to the continuous time Markov process; the major difference is that the holding function may be of any form, and different for any i and j. When the $h_{ij}(m)$ are exponential and independent of j, it reduces to the continuous time Markov process (Ginsberg, 1971). The model can also be derived from renewal models (Cox and Miller, 1965), as it is a cross between these and Markov processes.

The utility of this type of model is that Markovity assumptions can be relaxed, but the model is still "mathematically tractable in the sense that explicit expressions for many aspects of the process can be derived analytically from the assumptions defining the process" (Ginsberg, 1971). This means that relatively few parameters are needed to gain a lot of information about a process, even when duration of residence effects are included. In this sense, it is a step beyond McGinnis' work, making his approach to the mobility problem more tractable. The type of information that may be gained includes: limiting probability vectors, mean first passage times, mean recurrence times, mean state occupancies, and renewal functions (giving the overall probability of entering any state j in t time periods given the initial state of the system). The derivation of such properties is possible because of the intimate relation between the properties of the embedded Markov chain and those of the semi-Markov process:

State i is recurrent iff state i is recurrent in the embedded Markov chain.

State i is transient iff state i is transient in the embedded Markov chain.

State i communicates with j iff state i communicates with j in the embedded Markov chain.

A class is closed iff it is closed in the embedded Markov chain.

Given a finite number of states, state i is ergodic iff i in the embedded Markov chain is ergodic, and the mean waiting time in state i is finite.

(paraphrased from Ginsberg, 1971; originally quoted from Pyke, 1961).

Two examples will be given of now properties of the Semi-Markov process can be seen to be analogous to those of the Markov process modified by the inclusion of the waiting function.

(i) The limiting vector.

The limiting vector Π for the embedded Markov chain (which exists as long as this chain is ergodic) is easily calculated.

Given this (which is of course also the limiting vector for any simple Markov process), then the equilibrium vector for the Semi-Markov process is:

$$\Pi'_j = \frac{\Pi_j m_j}{\sum_{\ell=1}^{k} \Pi_\ell m_\ell} \qquad (j = 1\ 2\ \ldots .\ k)$$

$$\text{Where,}\quad m_j = \sum_t t.\ W_i(t)$$

m_j is the mean time to leave state i the mean waiting time.

(ii) The mean first passage time μ_{ij}.

This is the expected time taken for the process to enter a state j for the first time, given the process was in state i initially.

For the embedded Markov process let $g_{ij}(t)$ be the probability of travelling from i to j for the first time in t time periods.

Then the mean first passage time is:

$$u_{ij} = \sum_t t \cdot g_{ij}(t)$$

this can be computed recursively from:

$$g_{ij}(t) = q_{ij}(o,t) - \sum_{m=1}^{t-1} g_{ij}(m)\, q_{jj}\,(t-m,t)$$

from which the mean can be derived. $g_{ij}^{(2)}$ can be calculated from $g_{ij}^{(1)} = P_{ij}$; $g_{ij}(3)$ can be obtained from $g_{ij}(2)$ and so on.

For the semi-Markov model, the Mean first passage time from i to j is:

$$\sum_t tg_{ij}(t) = u_{ij}$$

where

$$g_{ij}(t) = q_{ij}(o,t) - \sum_{m=1}^{t-1} g_{ij}(m) \left[W_j(t-m) + P_{jj} . h'_{jj}(t-m) \right]$$

This expression can be seen to be similar to that for the simple Markov process (as given for the embedded chain), weighted by the functions which specify the probabilities of movement within a certain period.

The data needs for the semi-Markov model are greater than for the Markov model, as not only are probabilities of change required, but also empirical distributions for the rates of change. These two components are difficult to obtain separately, because empirical data on the number of people moving between states depends on both the atemporal probability of such a movement and on the number likely to move in a given time (h_{ij}). In any real data set these two factors may be both simultaneously affecting the observed frequencies. It might indeed be better to work the other way; trying out alternate holding functions, perhaps theoretically derived, and testing the model's sensitivity to variations in these.

However if data or theories are available, the model can be recommended as a way of relaxing some limitations of the simple Markov model without losing all of its computational neatness. Models have been proposed for migration (Ginsberg, 1971, 1972a, 1972b: a model which has the disadvantage of considering only one cohort of a common age at a time), housing turnover (Gilbert, 1972), and land use change (Drewett, 1969). None have yet been operationalized partly due to data difficulties.

Certainly the simple Markov model is not obsolete; for processes where duration of residence is not a dominant factor, the semi-Markov model would lose many of its advantages.

3.5. Inhomogeneous Markov Processes

Very few ongoing empirical processes, when modelled as a Markov process, tend to be found in anything like the equilibrium situation one might expect if a homogeneous process really had been operating for a period of time. This raises the question of the validity of the concept of stationarity. Undoubtedly, when probabilities are allowed to vary, computational difficulties increase as well as calibration problems, and the elegant conceptual framework of the homogeneous model begins to break down. There is no general theory of the behaviour of inhomogeneous Markov processes now, and nor is one likely to develop in the near future. The question really is whether it is possible to modify the concept of stationarity, and gain an increase in our insight into real world processes, without losing at the same time too much of the clarity of the model.

To give the concept of stationarity any meaning at all, variation in some fairly systematic way with respect to time or other independent variables should be attempted, at least initially. Otherwise there would be no way of predicting behaviour in the model, making modelling and forecasting for such a process quite futile. Gale (1972) has reviewed the problem of modifying the concept of stationarity, defining three types:

(i) Local Stationarity.

Under this, the present state of the process depends on its immediate past history, and the parameters determining such dependence are seen as being independent of the time at which they are measured, remaining constant under all other influences.

(ii) Functional Stationarity.

Here stationarity is defined in terms other than with respect to time. If the P_{ij}'s were determined by a model of the form:

$$P_{ij} = a + b\,x_i \qquad \text{(a and b are parameters)}$$

then a and b are seen to remain constant over the entire time period. Thus if x_i, say, were to change over time, the P_{ij}'s would change; meaning that local stationarity in the process (homogeneity) would be lost, while this functional stationarity would still hold. Development of this approach to modelling needs the linking of theory to the variables in Markov models, as envisaged by Ginsberg (1972a, 1972b), and Spilerman (1972b) but not yet operationalized. For this reason, this type of stationarity will not be discussed further here.

(iii) Differential Stationarity.

This is described as ". . . a means for discerning stationarity in particular kinds of non-stationary processes" (Gale, 1972). Here it is meant that if there is a process in which the P_{ij}'s are changing through time and this change is regular, then the difference between the probabilities at successive time periods is constant and differential stationarity exists.

For example,

$$P_{ij}^{(t+1)} = q\,P_{ij}^{(t)} \qquad \text{for all } t$$

where $P_{ij}^{(t)}$ is the probability of moving from i to j in time period t to t+1.

In what follows, the general structure of inhomogeneous processes will first be described, and then methods of getting round some of the difficulties in order to make the model more tractable (including use of

the concept of differential stationarity) will be outlined.

3.5.1. The General Mathematical Problem

As a comparison with the other forms of Markov processes it will be useful first of all to derive the C-K equations for the inhomogeneous model. For this the P_{ij}'s must be redefined, as they can be different for each time period. So $P_{ij}^{(t)}$ is to be used as defined in the example above. Then we may define the multi-step transition probability matrix as:

$$Q(t,T) = Q(t,r).Q(r,T) \qquad \text{for any } r: \quad t < r < T$$

For homogeneous processes, only the size of the time interval (T-t) matters, but here the actual values of both t and T matter as the one-step probability matrix P(t) depends on these.

Then the forward C-K equation would be (Howard, 1971):

$$Q(t,T) = Q(t,T-1)\cdot P(T-1)$$

and the backward C-K equation,

$$Q(t,T) = P(t)\cdot Q(t+1,T)$$

where in general $P(T-1) \neq P(t)$, and $Q(t, T-1) \neq Q(t+1,T)$.

Recursive solution of either of these would give:

$$Q(t,T' = P(t)\cdot P(t+1)\cdot P(t+1) \quad \ldots\ldots\ldots\ldots \quad P(T-1).$$

To use this, a means of determining the P_{ij} values separately for each time period would be necessary.

The difficulty with such processes arises because long-term and limiting properties of behaviour become far more complex to calculate. Howard (1971) gives an example of this. Given a process where the transition matrix is,

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-p^t & p^t \\ 0 & 1 \end{pmatrix} \end{array}$$

(implying the probability of entering state 2 decreases with time) then its behaviour can be compared to that of the homogeneous process with transition matrix,

$$\begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}$$

Then the limiting probability of being in state one as t in the latter process becomes:

$$(1-p)^t = (1/2)\cdot(1/4)\cdot(1/8)\cdot(1/16)\cdot(1/32)\cdot\,.\,.\,\underset{\text{as } t\to\infty}{\to 0} \quad \text{when } p = 1/2.$$

However, for the former process, with p also equal to ½, the series is:

$$(1-p).(1-p^2).(1-p^3)\;.\;.\;.\;.\;(1-p^t) = \prod_{i=1}^{t} (1-p^i)$$

$$= (1/2)\cdot(3/4)\cdot(7/8)\cdot(15/16)\;.\;.\;.\;.$$

Here the terms are less than one, but are increasing rather than decreasing, and the solution is more difficult. In fact only upper and lower limits to the limiting probability are obtainable.

Other features of the simple process also become far less tractable. For instance mean first passage times and recurrence times depend on the time t the process is in. It is conceivable that parameters like these could be solved for as values which are a function of time, but it is worth considering whether the gain in information would really compensate for the extra work necessary.

Hajnal (1956) has developed some general conditions for the ergodicity of inhomogeneous Markov processes. He distinguishes between <u>strong ergodicity</u> where the process tends to a limit independent of initial conditions, and

weak ergodicity where the system becomes independent of initial conditions (giving a matrix with identical rows) but does not tend to a limit.

Also Hajnal gives some conditions under which a matrix will not be ergodic and will instead progress to a periodic system, in which the process continually cycles around a group of states, or a decomposable system, in which the process breaks down into more than one exclusive intercommunicating set of states.

The concept of weak ergodicity is interesting in that it illustrates the fact that a process may be such that it is independent of initial conditions, but never stops changing; implying that some form of truly dynamic equilibrium may be in operation. In homogeneous processes only static types of equilibrium are possible. This illustrates the type of gain in conceptual content of models of empirical systems that is possible with inhomogeneous processes, but is gained at the expense of being able to predict the time paths of processes only under certain conditions.

3.5.2. Utilizing Differential Stationarity

An approach to dealing with this special case of inhomogeneity is proposed by Harary et al. (1970) who explicitly examine the differences between one-step transition matrices. They define a *differential matrix* as existing such that the P(t) are related according to the form:

$$P(t+1) = P(t)\cdot C(t)$$

where C(t) is the differential matrix, then $P(2) = P(1)\ C(1)$, $P(3) = P(2)\cdot C(2)$, and so on. The C(t) can then be calculated from:

$$C(t) = P(t)^{-1}\cdot P(t+1)$$

Then if all the C(t) were the identity matrix I, this would be a homogeneous process.

Harary et al. consider in detail the special case of this where differential stationarity exists such that:

$$C(1) = C(2) = C(3) = \ldots\ldots\ldots = C(n) = C$$

then,

$$C = P_{(1)}^{-1} \cdot P_{(2)}^{-1} \cdot P_{(3)} = \ldots\ldots P_{(t-1)}^{-1} \cdot P_{(t)}$$

This also means,

$$P_{(2)} = P_{(1)} \cdot C$$

and,

$$P_{(3)} = P_{(2)} \cdot C = P_{(1)} \cdot C \cdot C$$

or in general;

$$P_{(n)} = P_{(1)} \cdot C^{n-1}$$

Then the multi-step transition probability matrix would be of the form:

$$Q(t,\ t+1) = P(t) = P_{(1)} \cdot C^{t-1}$$

Therefore,

$$Q(t,T) = Q(t,\ t+1) \cdot Q(t,\ t+1) \cdot C \cdot Q(t,\ t+1) \cdot C^2 \ldots.$$

$$= \prod_{s=t}^{T-1} (P_{(1)} \cdot C^s)$$

Then it can be shown that under certain conditions this tends to a limit as $n \to \infty$ Whenever C is stochastic, then so is $\prod_{s=o}^{\infty} (P_{(1)} \cdot C^s)$, although this latter may contain negative elements. Results have only been gained for a two state process.

It is not the detailed results that matter so much here as the method, which allows statements to be made relatively free of conditions giving specified behaviour of this type of process. These results are general enough to allow the delimitation of conditions defining when C and $P(1) \cdot C^s$ are and are not stochastic.

(Harary et al., 1970). It is a form of calculus of differences, equivalent to taking differentials. The result is that analysis is concentrated on rates of change rather than actual change. When the rate of change is constant, then it becomes possible to examine the limiting behaviour of the constant differential matrix, even when this cannot be done for the continually varying P(t) matrix. Thus the overall analysis of the process gains more power.

It is also possible to think in terms of second differences equivalent to a second differential, where for instance:

$$C(t+1) = C(t) \cdot E(t)$$

This in effect would allow analysis of situations where the rate of acceleration in the rate of change of the P(t) matrices is constant. It is also worth noting that all these differences have been taken on the right hand side. Because matrix multiplication is not in general commutative, left hand differences of the form,

$$P(2) = D(1) . P(1)$$

should also be considered.

> This analysis has provided interesting insights into real-world phenomena that have been empirically validated. It has effectively provided the basis for powerful models of previously intractable solutions. From a mathematical point of view, the work opens a new domain for a probabilistic investigation of chains. (Harary et al., 1970)

Whether these claims turn out to be validated depends on future work and generalization away from the two state case, but it certainly provides an attractive alternative approach to some very sticky mathematical problems.

3.5.3. Partially Homogeneous Processes

Utilizing the methods of statistical analysis of Markov processes, Henry (1971) provides a more empirical approach to simplifying the problem

of inhomogeneity. This involves breaking the transition matrix into two parts and analyzing them separately, using empirical data to test for stationarity. If only part of the matrix is subject to change through time, then it may be easier to say something about its long term behaviour.

Henry takes the case where the main diagonal of the matrix, representing the probability of migrants not moving, may vary through time; whereas the probabilities governing their choice of states to move to (the off-diagonal elements) may not vary. Therefore it is desirable to separate these two elements.

To do this, two matrices are defined:

D(t), which contains just the diagonal elements of the empirical transition probability matrix at time t;

M (t), which is the rest of the transition probability matrix at time t (i.e. zeros on the main diagonal), but with the elements on each row weighted up so that M(t) is stochastic (rows summing to one).

Then it can be shown that:

$$P(t) = D(t) + (I-D(t)) \cdot M(t)$$

Now from empirical data, the maximum likelihood estimates of the probabilities for D(t) and M(t) are obtained for each time period. Also calculated are the probabilities that would occur if D(t) and M(t) did not vary over time, i.e. if:

$$D(1) = D(2) = \ . \ . \ . \ . \ . \ . = D(T) = D$$

and $$M(1) = M(2) = \ . \ . \ . \ . \ . \ . = M(T) = M$$

D is calculated as the average value of all the D(t) matrices (t 1, 2, ..., T), and similarly for M.

Then three hypotheses are tested:

HO: D and M do not change through time.

H1: D changes but M does not.

H2: Both D and M change through time.

These hypotheses are tested in a nested form; H1 is tested against H0, and H2 against H1.

For example for a χ^2 test we can define:

$$P^{(0)}_{ijt} = \hat{d}_{ij} + (1 - \hat{d}_{ii})\, \hat{m}_{ij}$$

$$P^{(1)}_{ijt} = \hat{d}_{ijt} + (1 - \hat{d}_{iit})\, \hat{m}_{ij}$$

$$P^{(2)}_{ijt} = \hat{d}_{ijt} + (1 - \hat{d}_{iit})\, \hat{m}_{ijt}$$

and then,

$$\chi^2_{(0,1)} = \sum_{t=1}^{T} \sum_{i=1}^{k} f_{it} \cdot \sum_{j} (P^{(1)}_{ijt} - P^{(0)}_{ijt})^2 / P^{(0)}_{ijt}$$

$$f_{it} = \sum_j f_{ijt}$$

H_o is the null hypothesis (T-1) d.f.

and similarly for

$\chi^2_{(1,\ 2)}$ at (T-1).k.(k-2) d of f.

In practice Henry found an insignificant result for $\chi^2_{1,2}$, and a significant one for $\chi^2_{o,1}$, implying that in their data set D was varying, but M was not.

Given this information, that much of the chain was stationary, it is possible to obtain results for the long-term behaviour of the process (Henry, 1971). If M is ergodic, and none of the entries in any of the D(t)'s increase or they all decrease over time; then it is shown that the ratio of the probabilities of leaving any two states, $\frac{(1 - P_{iit})}{(1 - P_{jjt})}$, must remain constant over time for the composite model to move towards a limiting vector.

Also the values in the matrix P(t) can be predicted:

$$P(t) = I - (c_t/c_1)(I-D(1))\ (I-M)$$

where $c_1, c_2, \ldots, c_t$ are a decreasing set of positive scalars representing the change in D(t) over time. Even if sequence of c_t's is not decreasing, the process will still move to a limiting vector as long as none of the c are so large with respect to c that the entries in (I-D(t)) are greater than one, since $(I-D(t) = (c_t/c_1).(I-D(1))$.

This model could be extended to a matrix M where, say, instead of separating out the main diagonal, it might be worth while examining whether the probability of entering a certain state j varies over time when the rest of the model does not. For instance migration into one city may vary due to external influences which do not affect the rest of the area. This would involve taking out the jth column from the P(t) matrices as a series of separate matrices G(t), analogous to the D(t); leaving behind the matrices M'(t) which have zeros in the jth column and like the M(t) are weighted up so that they are stochastic.

Then:

$$P(t) = G(t) + (I-D'(t))\ \ M'(t)$$

Note that the operator D'(t) is a diagonal matrix as before where the element $\{d_{ii}\}$ is equal to the element in the ith row of G(t). It is possible to generalize further by taking out any set of elements from the original probability matrix and putting them in a separate matrix G'. The above formula would still hold providing the d_{ii}th element in D'(t) is equal to the sum of all the elements in the ith row of G', and providing that none of the rows of M'(t) are left empty, as it would then no longer be stochastic.

Then providing M' is still ergodic and stationary (it must have no elements equal to one), Henry's arguments about conditions for a limiting

vector would still hold.

It should be remembered that there ought to be some theoretical reason for dividing the probability matrix P(t) into two parts. Systematically searching through P(t), until all the elements that are not varying through time in the empirical data are separated from those that do vary, does not achieve very much. If there is no real reason why certain elements should not vary, it might only be by chance that they are stationary in this data set; and treating them as if they were always stationary would lead to erroneous prections.

In this section, three very different approaches to the problem of homogeneity have been outlined: the traditional mathematical approach, an approach looking at rates of change of probabilities, and an attempt to pick out empirically the non-stationary element in a process which may be only partially inhomogeneous. It should be noted that none of the approaches has been taken very far, and there is no common concensus about aims and methods when dealing with inhomogeneous processes. In this paper emphasis has been put on conditions for the existence of a limiting vector and attempts to specify these. It should be realized that equilibrium vectors are not everything; just because they cannot be found it does not mean that the Markov model loses its validity. However the ease of dealing with this property is an indicator of the general tractability of the model, and this ability of the model to provide a lot of internal information based on relatively little data input should not be lost. As long as we have a way of specifying how the probabilities vary over time it will always be possible to calculate other properties of the process. Here again the need to link substantive theory to the empirical probabilities of the Markov model stands out as a major problem.

Analysis of inhomogeneous processes means that the Markov model is no longer a neat way of picking out the properties of a process. However such models do allow a relaxation of an assumption that may have been very restrictive on our model construction to date. To discover whether we really can drop this assumption and still use Markovian models, empirical evaluation of their utility is necessary.

3.6. Concluding Remarks

From the tremendous growth in interest in the types of subjects discussed in this paper, it is probably safe to conclude that the days of dominance of the simple Markov model are passing. In its place, a whole series of modified Markov models are being developed. This also means that data requirements are no longer restricted to a single group of figures from one time period, and as data availability never expands to fill the needs of researchers, much effort is going to be devoted to ways of getting around this problem. The work by Spilerman (1972b) and Ginsberg (1972b) in this field is likely to be the tip of an iceberg with obvious benefits for all kinds of stochastic model calibration.

As far as the development of the new models goes, it is worthwhile reviewing the question of better models versus difficulty in operationalizing them. The two major aspects dealt with here are the problems of relaxing Markovity and inhomogeneity. McFarland (1970) refers to the Markovity assumption as a form of scientific determinism, where the past entirely determines the future. This is a fairly accurate assessment of the situation and drawbacks underlying this approach. Few would argue that this is a very simple way of modelling process, but on the other hand the model possesses an ability to provide information unequalled by other

models of equivalent ease of operation - an elegance paid for by restrictive assumptions. Because of the undoubted utility of Markov models in research, many attempts to soften these assumptions within the Markov framework have been made, but with mixed success and usually an explosion in data needs. However there does exist a viable alternative - the semi-Markov model. Here by use of more sophisticated mathematical techniques nearly as much information can be gleaned from the model as from the Markov model, and although data needs are greater they are not wildly so. It is likely that this model has a considerable future in empirical research.

Unfortunately the semi-Markov model still assumes stationarity; and this is a problem which is not so easy to overcome. At present there is no ready mathematical theory that can be used to avoid this assumption, but this does not mean that the problem should be ignored.

Given that there is a need for this type of model, and that there is little mathematical theory to draw on; then it may be better to approach the problem from the other side, developing models for empirical cases as they arise. If there is a way of specifying how the probabilities change in certain types of empirical situations, then certain classes of inhomogeneous models that are of particular use in practice may become evident. If this situation arises then techniques, such as those developed by Harary et al. (1970) and Henry (1971), may reveal a great deal when effort is concentrated on these sub-groups. In particular situations a technique that may help reveal a lot about the behaviour of the process under study is Monte Carlo simulation. Simulation has been resorted to far too often as a sort of cure-all, and has been badly used at that; but careful construction and analysis of different models in this way may give rise to observable regularities in the behaviour of inhomogeneous processes.

To test the relative utility of these types of models and alternative ways of modelling the real world, needs above all empirical evaluation. No final conclusions can be drawn in a theoretical review of this type; all that can be done is to set out the pros and cons. One can conceive of new theoretical developments, such as linking inhomogeneous and semi-Markov models to kill two assumptions at once; but there is a danger in letting theory get too far ahead of practical application since all sight may be lost of the needs of empirical workers. Instead, much effort should be put into making the present methods applicable in practice, and ensuring that they are used with a proper awareness for the assumptions involved. Then a true idea of what is needed and what is possible can be obtained. It may turn out that only the simple Markov model is practicable, and that there are better techniques for doing the job of inhomogeneous and Semi-Markov models; but the only way to discover this is by attempting to use all three in practice.

CHAPTER IV

SOME FUNDAMENTAL APPROACHES TO MARKOVIAN MODELS OF URBAN CHANGE

4.1. Introduction

For the most part, the Markov chain applications thus far considered have been heavily empirical. That is, most of the models have used relative frequencies of transitions as estimates of the true transition probabilities. These methods therefore often have excessive data requirements. Thus for an N state Markov chain, N^2-N parameters must be estimated. Since the number of states is frequently quite large, particularly in spatial models, these data needs can be quite formidable. Moreover, a rationale for using a Markov chain model is frequently not clear and often not even given. In the light of these comments, there appears to be a need to develop models with a sounder theoretical rationale to obtain predictions and/or insights in which one can have a greater degree of confidence and also to reduce the large data needs of some Markov formulations.

In this section then some basic probability models are considered and some suggestions are made about their apparent relevance to urban problems. These models include random walks, birth and death processes, queuing theory, and diffusion theory. In addition, deterministic models with a Markovian lag structure are considered. Much of the material is adapted from Bartholomew (1967), Tintner and Sengupta (1972) and basic texts in operations research, notably Wagner (1969) and Hillier and Liebermann (1967).

4.2. Random Walks, Birth and Death Processes

Consider any urban phenomena which can either increase or decrease in size as time passes. The state of the process in question could then refer to the number of jobs, dwelling units, retail establishments or other entities of interest to a decision maker. To fix our ideas, we shall use the example of population of the urban area in the following discussion.

Let us first assume the city's population may increase by one unit, remain the same, or decrease by one unit over a unit time interval. We can for many practical purposes exclude the possibility of larger jumps in the population by making the time interval sufficiently small. By this device, we will have, by definition, a very sparse transition probability matrix with the positive probabilities arrayed along the principal diagonal. If we assume the probabilities of transitions to neighbouring states are independent of the current state (except perhaps for boundary states), then the Markov process is a random walk. If the probabilities can vary from one state to another so that, for example, increases are more likely for a large city than a small one, the model is referred to as a birth and death process.

Matrix A is an example of such a Markov process.

$$A = \begin{pmatrix} 1-b_o & b_o & 0 & \cdots & & \\ d_1 & 1-d_1-b_1 & b_1 & 0\ldots & & \\ & & & & & \\ & \cdots & 0 & d_{N-1} & 1-b_{N-1}-d_{N-1} & b_{N-1} \\ & & & \cdots & 0 \; d_N & 1-d_N \end{pmatrix}$$

This is a finite state birth and death process where the city's size is limited to integral values between 0 and N. For such a model, in addition

to all of the theorems of finite Markov chains, there are some additional, more general, theorems arising from the rather simple structure of the probability matrix. For example, general solutions for the limiting state probabilities are available (see Howard, 1972, Ch. 7). Even in the semi-infinite case where the city's population is not limited to a maximum value of N such general results are available if b_k the probability of a unit increase is less than .5.

Some of the models of this type have been developed to study queuing systems whereby items arrive to be serviced according to a certain probabilistic rule and are serviced or leave the queue according to another rule. The simplest process of this kind is a linear, time-invariant, pure birth process. In an urban growth context, a city of size X can either remain at size X or increase to size X + U with probability $\lambda(\Delta t)$ in the time interval t, t + Δt. Other changes can occur with probability $0(\Delta t)$, but these are assumed to be negligible. That this is a Markovian process can be readily seen from the following basic probabilistic equation:
$P_x(t + \Delta t) = P_x(t).(1 - \lambda\Delta t) + P_{x-u}(t).\ \lambda\Delta t$ where $P_x(t)$ is the probability the city will be of size x at time t. Using these quite primitive assumptions, it is easy to derive the time dependent probability distribution of the city's population:

$$P_x(t) = \frac{e^{-\lambda t}(\lambda t)^{x/u}}{(x/u)!}$$

This is a modified Poisson distribution with mean $(M(t) = u\,\lambda t$ and variance $V(t) = u^2\,\lambda t$. This very simple growth pattern, a homogeneous linear trend, is shown in Figure 4.1.

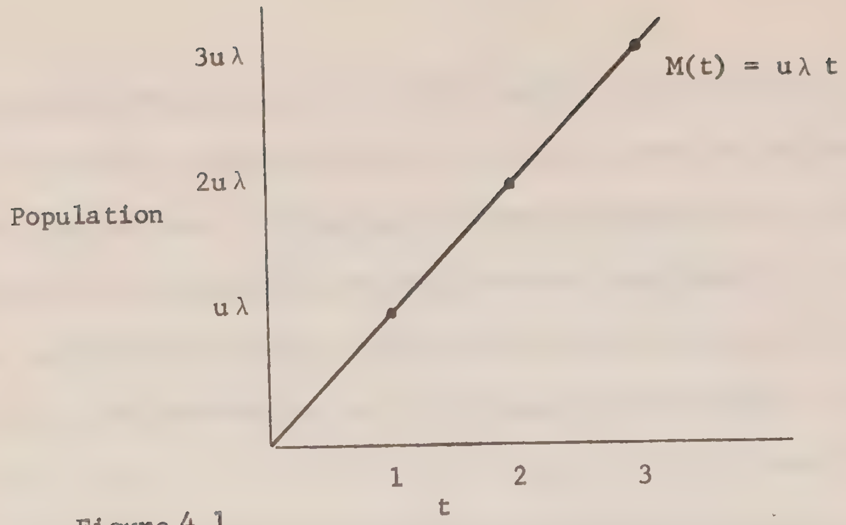


Figure 4.1.

The variance also increases linearly with time, but much more rapidly than the mean. It is understandable that considerable imprecision is associated with such a simple model.

A very minor modification allows for a positive population at time 0, and a linear inhomogeneous expected trend is derived.

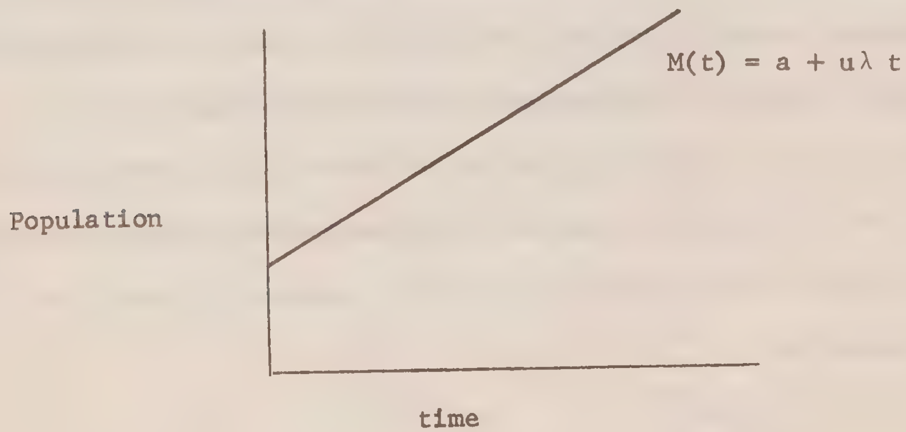


Figure 4.2.

Note that neither u, λ, or in the second formulation a, need be known <u>a priori</u>. Using well known methods of least squares, these parameters may be estimated from the appropriate time series.

This probabilistic derivation gives some <u>a priori</u> justification or

rationale for a linear trend and also gives the analyst a clear statement of the expected deviation from this trend, an extremely important feature when dealing with systems which are volatile and not fully understood. In addition to the moments, the actual probability distribution is derived so that expected values of costs and benefits associated with various population levels could be calculated.

A slight mathematical generalization which is extremely important substantively is to make the birth rate λ time variant. This too results in a modified Poisson of the following form:

$$P_x(t) = \frac{\exp\left(-\int_o^t X(Z)\, dZ\right)\left(\int_o^t \lambda(Z)\, dZ\right)^{\frac{(x-a)}{u}}}{[(x-a)/u]\,!}$$

with mean $M(t) = a + u \int_o^t \lambda(Z)\, d\,Z$

and variance $V(t) = u^2 \int_o^t \lambda(Z)\, d\,Z$

Making the birth rate λ a function of expected trend in per capita income, the size tertiary sector, or perhaps a policy variable such as government expenditures in the urban centre, quite complicated non-linear trends could be generated. This strategy of linking a simple probabilistic model with a deterministic model by making the parameters of the first a function of the behaviour of the second can be a productive yet simple way of developing quite realistic models.

Perhaps a more interesting case of the pure birth process is where the birth parameter is a function of the size of the current population. Thus for example, larger towns tend to have larger growth rates for a variety of reasons. This model differs from the preceeding models in that the parameter value change is endogenous to the model not determined arbitrarily or by an associated deterministic model.

The birth parameter can be assumed to be proportional to town size $\lambda(t) = \lambda x(t)$. This assumption yields an exponential trend $M(t) = j \exp(\lambda t)$ with probability distribution $P_x(t) = \begin{pmatrix} x-1 \\ x-j \end{pmatrix} \exp(-j \lambda t)(1-\exp(-\lambda t))^{x-j}$ where $j = x(o)$ and variance $V(t) = j \exp(\lambda t)(2j \exp(\lambda t)-1)$.

These results can be generalized to a mixed birth and death process. In this case, there is also a certain probability $\mu x(t)\Delta t$ that the city will decline by some quantity u in the time interval $t, t + \Delta t$. This could result for a variety of reasons including congestion diseconomies of large urban areas. Where $\mu(t)$ and $\lambda(t)$ are both constant proportions of $x(t)$ it is a simple matter to define a composite parameter $\bar{\lambda} = \lambda - \mu$ as the net birth rate. This yields the same probability distribution and exponential trend as the previous model but some adjustments are necessary for the higher moments. The possibility of movements in both directions obviously implies, for example, that variance is larger for the second model.

If the constants of proportionality λ and μ vary over time Tintner and Sengupta (1972) suggest that the system's behaviour can be partitioned into regimes or periods within which both μ and λ are constant (or approximately so). Expected growth or decline in the k^{th} regime is governed by the current value of $\bar{\lambda}$. Interestingly it can be shown that the long run trend $M(t)$ is entirely determined by the overall <u>mean</u> values of μ and λ. If the regimes are of equal size, for example, a mean birth rate greater than (or less than) the death rate would indicate that the long run behaviour of urban growth (or decline would follow an exponential time path. Superimposed on this long run growth trend could be considerable fluctuations in expected town size reflected the shifts in $\bar{\lambda}_k$. Figure 4.3 illustrates one possible path of urban population change which could be generated by such a model.

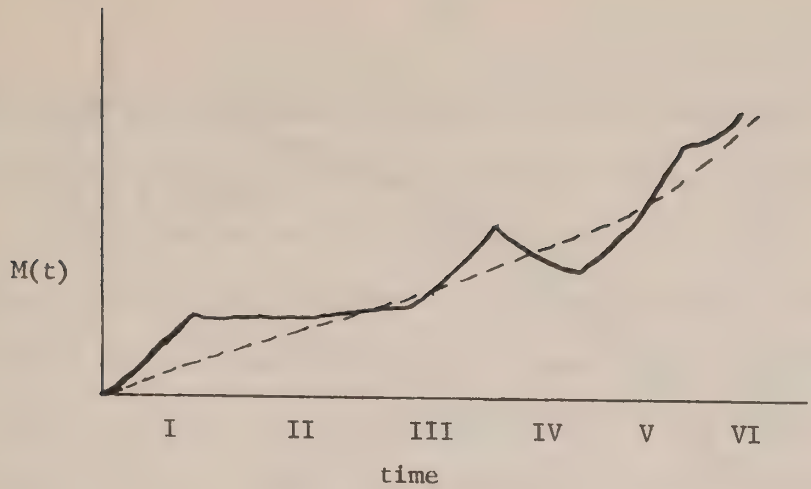


Figure 4.3.

An extension of some interest in one in which one of the parameters, say μ_x, is proportional to the current value of x (i.e., $u_x = cx(t)$), but the other is not ($\lambda x = ac$) where a and c are positive constants. The resulting probability distribution is Poisson

$$P_x(t) = \frac{\exp(-h(t))\,(h(t))^x}{x!}$$

with mean and variance $M(t) = V(t) = h(t) = a-b \exp(-ct)$ where

$b = a - j$ and $j = h(o)$. Thus in this model the growth is retarded by the "death rate" to an increasing extent as the process proceeds. In the initial stages, growth is rapid but declines linearly with larger populations ($\frac{dx}{dt} = ac - cx$) nearing a zero rate of growth as the expected asymptotic population level is approached. Figure 4.4 illustrates the type of trend generated if $h(o) < a$. If $h(o) > a$, an exponential decline asymptotic to a would result.

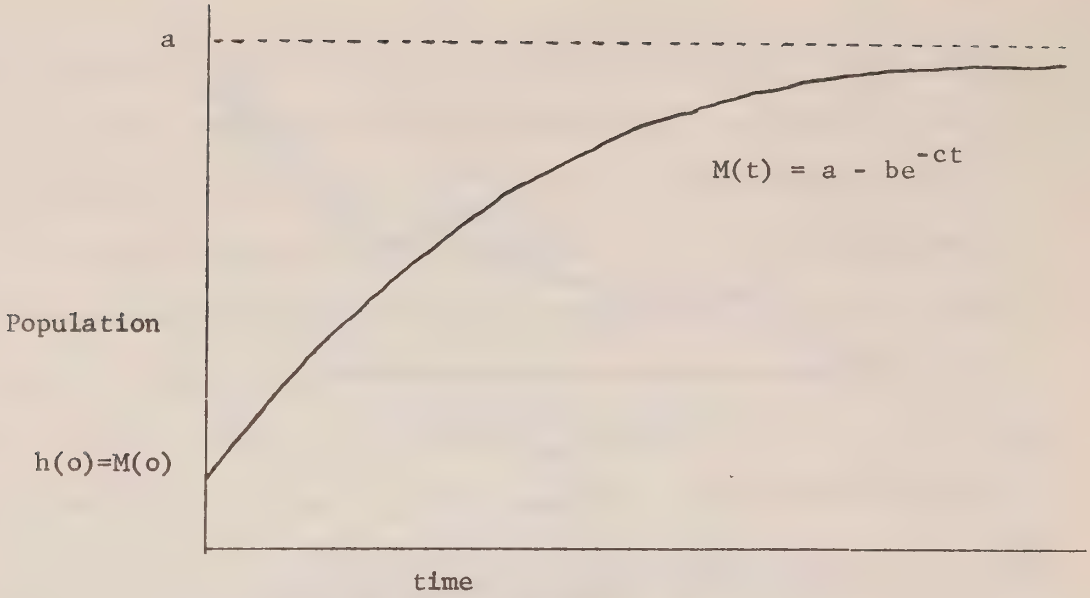


Figure 4.4.

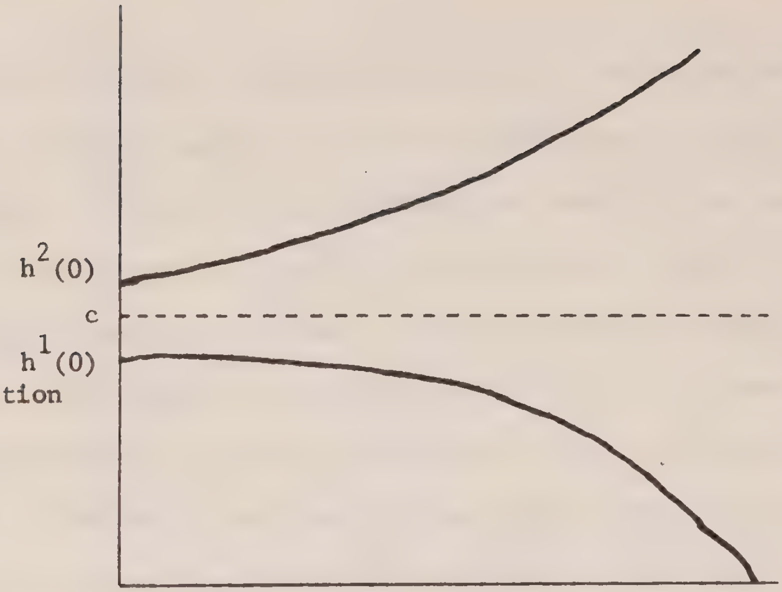
Tintner and Sengupta (1972) emphasize another characteristic of this model. The limiting behaviour of the Poisson probability distribution is $\lim_{t \to \infty} P(t) = \frac{\exp(-a)\ (a)^x}{x!}$. That is, the long run probabilities of state occupancy are independent of the initial state of the system. Recall that the limiting behaviour of certain types of Markov chains have this characteristic. This birth-death process is also based on Markovian principles, hence the similar results. These findings anticipate and underline the major thrust of the arguments of Oates, Howrey, and Baumol (1972) and Baumol (1973) in their deterministic difference equation models of urban change. Altering the initial conditions (variously referred to as j, h(o), x(o)), for example, ensuring by one means or another that the level of population (or housing stock, etc.) at time zero is some value, say Q, will in no way affect the limiting probability distribution. To change the asymptote, the parameters of the process (c and/or a itself) must be altered. These parameters reflect, of course, the relative size of the probabilities of growth and decline at

various stages in the urban development process, probabilities which are related to social and economic processes and perhaps government decisions.

Another feature which distinguishes this model from previously discussed birth-death processes is that it does in fact have an equilibrium expected population level. Some would consider simple exponential growth models to be less than satisfactory in that they yield trends which go to infinity, clearly an impossible result in most situations. This criticism, however, is not terribly serious if one recalls first the aphorism of Keynes that "in the long run, we are all dead" and secondly that long before long run behaviour is encountered in most social contexts, the parameters of the process change and a new expected trend comes into effect.

In this connection, also note that the variance about this expected trend is exactly equal to the trend itself (i.e. $V(t) = M(t)$). Thus, the expected precision of the model's predictions is not very great for larger populations. These long run equilibrium properties should therefore be viewed with interest but tempered with considerable caution even in the case where parameters are constant.

Alternatively and perhaps somewhat more realistically in the urban growth context we assume the "birth" rate is proportional to city size and the "death " rate is constant we obtain an entire set of analogous but different results. The expected trend for example would be one of the two types shown in Figure 4.5 depending upon whether $h(o) < c$ or $h(o) > c$. The latter yields more plausible results in that it tends to $+\infty$ while the former tends to $-\infty$. Of course, in these cases no equivalent steady state probabilities exist.



M(t) = C - (c-h(o)) exp (at) for two different assumptions regarding h(o)

Figure 4.5.

Other probabilistic processes with different trends may be derived by assuming for example $\lambda(t) = \lambda_o + \lambda_1 x(t)$ and $u(t) = \mu_o + \mu_1 x(t)$ (Tintner and Sengupta, 1972).

4.3. Queuing Theory Models

Queuing theory is concerned with the analysis of servicing problems where units to be serviced are arriving according to some "birth" process and leave the queuing system (i.e. they are given service - tune-ups, hamburgers, court trials, etc.) according to a specified "death" process. Queuing theory then deals with a special case of birth and death process. We discuss it separately from the more general processes because it is extremely well developed with its own literature and jargon and has formulated and solved variants of birth and death processes some of which may be relevant

to the modelling of urban change. Using the analogy between size of the queuing system and city size, expected waiting time and mean length of residency, it is possible to develop many alternative urban models. We have already seen, for example, that for a pure birth process with a constant birth rate λ, the probability distribution of city (queue size is Poisson with mean and variance λt. It can also be shown that for a pure death process with constant death rate μ, the probability distribution of city (queue) size is truncated Poisson with mean and variance μt.

Queuing theory is concerned with deriving certain statistics (expected queue length, expected waiting times, etc.) arising from systems which are characterized by these and other input and output distributions. For both substantive and analytical reasons, steady state behaviour is often emphasized. The time it would take a machine shop for example, to reach a steady state would be small relative to the life of the facility; thus, steady state behaviour is highly relevant. For the system to reach a steady state, it is necessary for the mean arrival rate λ to be less than the mean service rate μ; otherwise the size of the queue would grow without bound - an untenable arrangement for any machine shop! That unbounded (e.g. exponential) growth is not quite as unrealistic in an urban context does not invalidate queuing theory results; it simply indicates that there is a broader range of potentially relevant parameter values in the urban case.

For the combined simple birth-death process above with $\lambda<\mu$, steady state probabilities can be easily derived:

$$P_n = P_o \left(\frac{\lambda}{\mu} \right)^n$$

where P_x is the probability that the size of the city (queue) will be X in equilibrium. Expected system size is then $\frac{\lambda}{\mu - \lambda}$.

Thus a village with mean in-migration plus birth rate of 100 per year and mean out-migration rate of 102 per year will have an expected population size of 50 in the steady state. The number of movements in and out is clearly not realistic in this example. One should recognize however, that the choice of time unit is arbitrary. The rates could be in terms of decades rather than years. The same equilibrium results would be generated.

A more promising tactic would be to allow arrival rates to exceed departure rates at least for some ranges of population size - a range over which economies of scale outweigh diseconomies. Many state dependent arrival and departure queuing models have been formulated and solved. Most assume monotonically decreasing birth rates and increasing death rates with queue length. For example, $\lambda_x = (X + 1)^{-b} \lambda_0$ and $\mu_x = X^c \mu_0$. In a dynamic city size model, while the second postulate may seem plausible, the first does not. Larger cities typically attract more jobs and people than smaller ones. It is possible however that after a certain size is reached, the city becomes less attractive so that arrival rates begin to decline and therefore become less than departure rates (see Figure 4.6.).

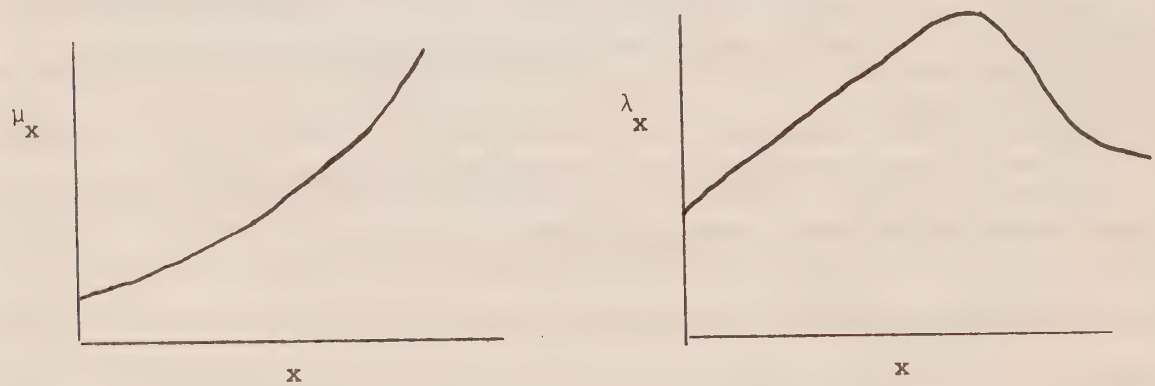


Figure 4.6. One assumption about Form of State - Dependent λ and μ.

These assumptions could take the explicit form of

$$\mu_x = x^c \mu_1 \qquad \text{and} \qquad \lambda_x = x^a e^{-\frac{x}{d}}$$

As long as there exists a state k such that $\mu_x > \lambda_x$ for all $x > k$, a steady state probability distribution of city size will exist. (This is a sufficient not a necessary condition.) Thus, an alternative postulate may be that both λ_n and μ_n increase monotonically with city size with μ_n convex and λ_n concave:

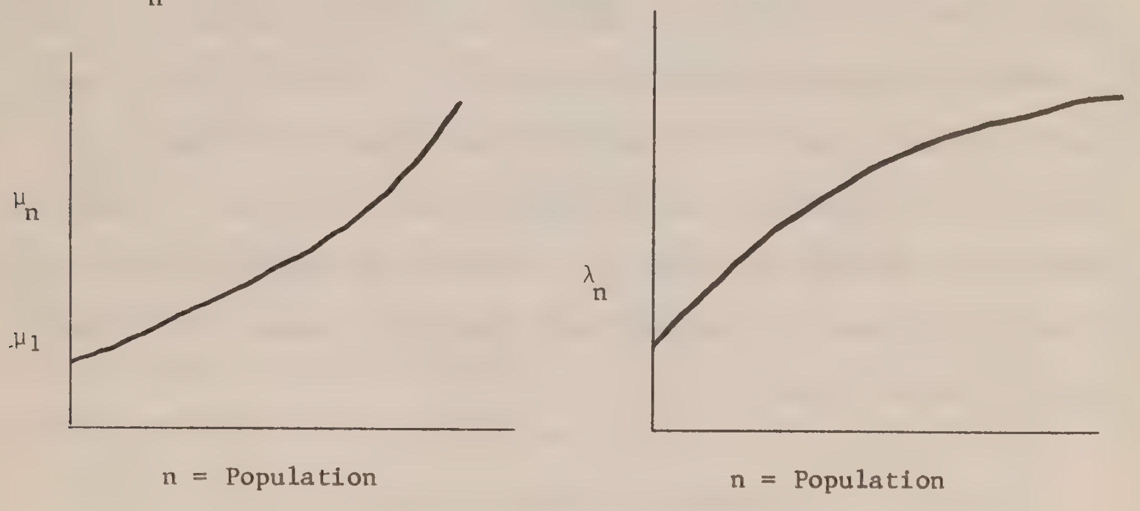


Figure 4.7. Alternative Postulated Forms of State Dependent Arrival and Departure Rates.

In addition to specifying the mean arrival and departure rates, one can postulate that arrivals are Poisson distributed with parameter λ_n and service times (length of residency in the urban case) are exponentially distributed[1] with parameter μ_n. This assumption conforms with the sociologists' "principle of cumulative inertia" (see McGinnis, 1968, Ginsberg, 1972b and discussion in Chapter III of this report).

---

[1] Assuming exponentially distributed residency times is equivalent to assuming that the number of departures over a given time interval is Poisson.

Making these assumptions, the steady state results for the model can be readily derived:

$$P_n = (n!)^{b-c} \left(\frac{\lambda_o}{\mu_1} \right)^n P_o$$

with P_o being estimated from the condition $\sum_{i=0}^{\infty} P_i = 1$. From the probability distribution, expected population size, expected residency times, and variance can be derived. The steady state probabilities for town size in this case are a function entirely of the difference between parameters b and c, by assumption a negative value, <u>and</u> the ratio of the basic arrival and departure rates λ_o and μ_1. It would be easy to generate a family of probability distributions, each relating to different values of these two quantitites. Holding the ratio λ_o/μ_1 constant, the probability distribution of town size becomes progressively more positively skewed as b-c takes on larger and larger negative values.

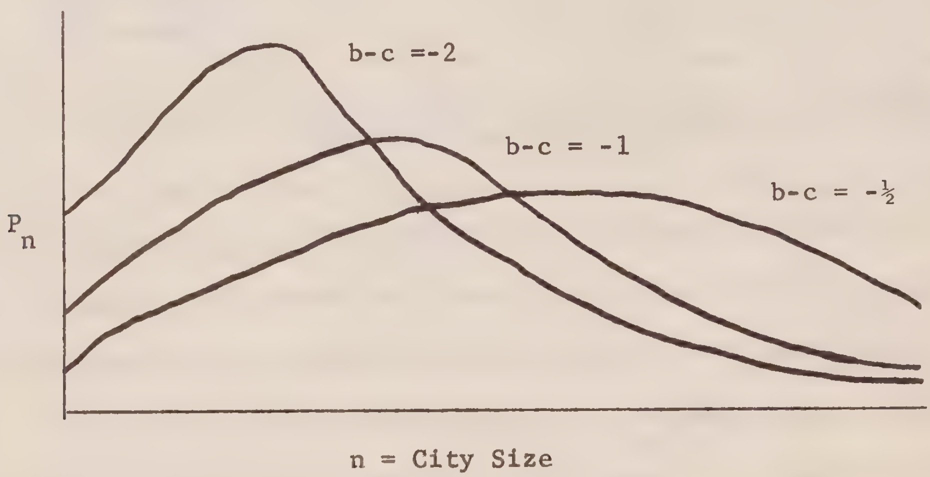


Figure 4.8. Illustration of the Relationship Between b-c and the Shape of the Probability Distribution of Population.

Similarly, as the ratio λ_o / μ increases, the curves become both more symmetric and centred around larger expected population levels.

Such models, derived from queuing theory, provide a stochastic framework within which urban growth processes may be studied. It may be possible to relate some or all parameters b, c, λ_o, and μ_1 to deterministic processes either of a descriptive or policy nature. As b approaches its upper limit of unity, the marginal attractiveness of increased city size decreases at a diminishing rate (i.e. larger cities become more attractive than with smaller values of b). Thus b could be directly related to some measure of comparative locational advantage of the town or alternatively to a technological environment which encourages or discourages larger cities. It is possible to envisage a two-tiered model, the first of which is deterministic, essentially setting the parameters for a system of cities; the second, stochastic, which determines the transient and steady state probability distributions for each of the cities.

One should note that the functional relationships for the state-varying parameters λ_n and μ_n need not be well-behaved or even in closed form. They may be quite arbitrarily discontinuous and non-monotonic incorporating such things as critical urban thresholds and ceilings, sharp policy reversals, etc.

Different results may be achieved if other probabilistic input or output mechanisms are postulated. One of the most frequently studied modifications assumes gamma (Erlang) distributed service or residency times. The gamma distribution is a very flexible non-negative distribution which can describe a wide range of mover behaviour. Although we have already noted an argument in the literature which would recommend the adoption of

exponentially distributed residency times (McGinnis, 1968; Ginsberg, 1972b), a gamma function may be more appropriate. Clearly it is rather improbable that a factory or a household will be more likely to re-locate one day after it has moved to a site than one year after. The shear costs of moving and the time it takes to become aware of the limitations of the new site would make such a postulate implausible. Early moves are thus unlikely. After a household has lived in an area a long time, and set up social and economic ties, the concept of cumulative inertia could begin to become effective. Thus one of the curves from Figure 4.9 could be selected as the probability density function of length of residency. Models with these types of service times have been studied extensively in the queuing theory literature.

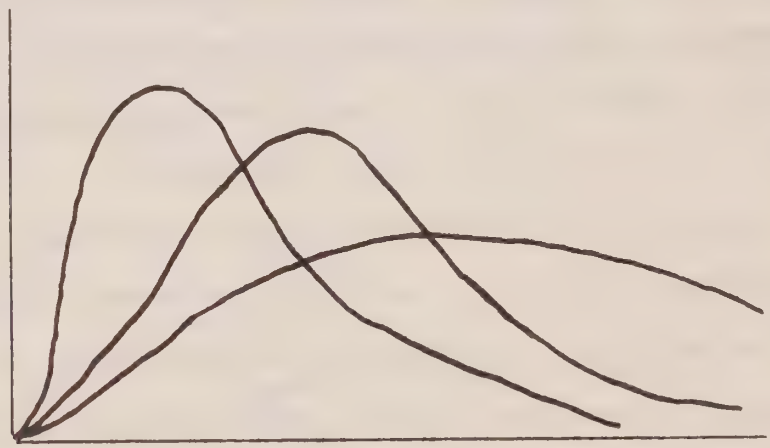


Figure 4.9. Examples of Gamma Distribution.

4.4. Multi-Regional Models

There appears to be no direct analogy in queuing theory which relates to the multiple city case of births, deaths, and intercity migration. Some of the findings can be used however when they are augmented with other models. The semi-Markov process models used by Ginsberg to describe migration represents such a synthesis. The time between moves in a location is exponentially or perhaps gamma distributed. The nature of the move is generated by the imbedded Markov chain. The arrivals are therefore a function of the departure rates and transition probabilities of the imbedded Markov chain.

This leads very naturally into the question of whether the intercity or interzonal transition probabilities can be based on some theoretical rationale. To reiterate an argument made earlier, this important for two reasons: first to engender more confidence in the results and secondly, to facilitate the estimation procedure. This second reason is particularly pertinent to a multicity or multiregional model. The number of parameters increases exponentially with the number of cities. Migration data are often in short supply, sometimes non-existent. Thus some knowledge about the causal mechanisms underlying the stochastic process would be invaluable as an aid not only in estimation but also for subsequent sensitivity analysis so important for policy considerations.

One approach is to postulate that the transition (migration) probabilities, are a function of one or more variables, characteristics of origin and destination zones, competing destinations, and perhaps the distance separating the zones. Such a procedure is used by Spilerman (1972b) with some success. Migration probabilities are assumed to be a linear function of "push" and "pull" factors and "resistance" between pairs of

zones. Multiple regression models are estimated. This approach relates the probabilistic process to a deterministic, causal one and in principle would enable someone to predict the migration and urban growth consequences (both transient and long run) of a change, in one or more of the exogenous variables - changes resulting either from a proposed policy decision or predictions from other models.

This approach, while a potentially useful addition to Markovian models of urban processes is not entirely satisfying. First, it has heavy additional data requirements, although it may substitute more readily available data for data which are difficult to obtain. Secondly, the theoretical justification for the set of independent variables and the functional form of the relationship is weak. The works of Ginsberg (1972b) and Cordey-Hayes (1972) represent attempts to replace this *ad hoc* empirical theorizing with something with a more fundamental basis. Of the two, it would appear that Cordey-Hayes is the more successful.

Ginsberg, dissatisfied with the lack of causal structure in Markov models, attempts to derive a theoretical structure of the migration process based on Luce's theory of rational choice. While the general explanatory structure is unassailable, the particular functional relationships developed are still derived from quite arbitrary *ad hoc* assumptions.

Cordey-Hayes (1972) and Cordey-Hayes and Gleave (1973) use an approach initially adopted in molecular physics. In an attempt to identify an underlying causal structure, they first decompose transition probabilities into two or more constituent probabilities. In the case of migration, the probability a_{ij} of a resident in zone i moving to zone j is written as the product of an "escape probability ε_i and a "capture probability" μ_j:

Imposing equilibrium balance conditions, the number of households in each zone can be written in terms of these probabilities

$$N_i = N \frac{\frac{\mu i}{\varepsilon i}}{\sum_j \frac{\mu j}{\varepsilon_j}}$$

where N is the total number of households within the system. The ε_i and μ_j can of course be considered as functions of the characteristics of the zones. The expected number of households in each zone in transient (non-equilibrium) stages is also expressed as a function of these escape and capture probabilities as well as the initial state of the system.

Continuing the physical analogy, Cordey-Hayes uses the continuous form of the transport equation:

$$\frac{\partial \eta x}{\partial t} = \frac{\partial Jx}{\partial x}$$

i.e. the rate of change of the number of entities per unit area at x is equal to the negative of the spatial gradient of the flow at x. Now Jx, the number of households crossing x per unit time is postulated to be equal to the product of ε, the intrinsic mobility of individuals (which is related to life cycle status and mean time between moves) and $\frac{\partial v}{\partial x}$, the spatial gradient of unemployment and income.

The Cordey-Hayes papers are preliminary and exploratory, pleas to include a stronger causal element in dynamic models, using kinetic theory of physics to indicate how this might be done. By no means is he trying to invalidate the Markovian approach - rather he is attempting to enrich it, relating the parameters of Markov models to underlying causal

variables. The later Cordey-Hayes and Gleave (1973) paper uses this theoretical framework in an empirical British migration context.

Multivariate and/or multicity stochastic processes are extremely relevant in an urban context. The behaviour of one variable (e.g. total disposable income) may affect the behaviour of another (retail sales) for the same urban area. Similarly the growth of one city can affect the growth of a nearby city either positively or negatively. In the two city case, the state of the system can be described as 1x2 vector (x_1, x_2) the defining probabilistic equation for a pure birth process is:

$$\text{Prob}(X\ (t+\Delta t) = x + \delta | X(t) = x) = \lambda_\delta(\Delta t) + 0\ (\Delta t)$$

where $\delta = (1,0)$, $(0,1)$ or $(1,1)$ and is a measure of the relative likelihood of each of these types of population increases, and $0\ (\Delta t)$ as previously is a negligible probability which can be subsequently ignored. The probability of no change is then

$$\text{Prob}(X\ (t+\Delta t) = x \mid X(t) = x) = 1 - \sum_{\delta \neq 0} \lambda_\delta + 0\ (\Delta t)$$

From these initial postulates, it can be shown (Tintner and Sengupta, 1972) that the marginal distribution of each component $x_i\ (t)$ has a Poisson distribution with mean and variance $\sum_\delta \delta_i\ \lambda_\delta\ t\ \ (= \lambda_{10} + \lambda_{11})\ t$ in the two city case.

To be less restrictive in terms of units of measurement and increase, Tintner and Sengupta (1972) suggest that the state variables be expressed as

$$Y_i(t) = a_i\ (t) + U_i\ (t)\ x_i\ (t) \qquad i = 1, 2, \ldots$$

where the parameters $a_i(t)$ and $U_i(t)$ may vary with time or be constants. This adjustment allows the increase in an interval Δt to vary from one point in the process to another maintaining a constant $\lambda_\delta\ (\Delta t)$ probability

of such change. The variables $Y_i(t)$ have a bivariate (multivariate in the general case). Poisson distribution with means and variances

$$M_i(t) = a_i(t) + U_i(t) \sum_\delta (\delta_i \lambda_\delta t) \qquad i = 1, 2, \ldots$$

$$V_i(t) = U_i^2(t) \sum_\delta (\delta_i \lambda_\delta t) \qquad i = 1, 2, \ldots$$

In addition, the covariances between the populations of the two cities is of interest:

$$\mathrm{Cov}\ (Y_1(t),\ Y_2(t)) = U_1(t) \cdot U_2(t) \sum_{1,2} (\delta_1 \delta_2 \lambda_\delta t)$$

For large values of t, the multivariate normal distribution may be used as a good approximation.

The parameters may be estimated using least squares procedures and marginal and joint confidence limits may be determined using the normal approximation. If $a_1(t)$ and $U_i(t)$ are constants then an expected linear trend is generated for each city. If the parameters can vary over time, then the resulting trend could be quite irregular.

4.5. Diffusion Processes: A Brief Note

To this point we have only considered models for which over time interval $(t,\ t + \Delta t)$ at most three alternatives were possible with non-negligible probabilities: (1) a decrease from state x to state x-u, (2) an increase from state x to state x+u or (3) no change in state. We now turn to an important generalization where an infinite array of changes is possible. To do this we assume the existence of transition probabilities $f(\tau, x; t, y) = P_r\ (X(t) = y \mid x(\tau) = x)$. The immediate objective is to derive the explicit form of this probability distribution. To do this, it is assumed that a mean change over the

interval $t + \Delta t$ is equal to $\lambda_x(t) = \lambda\ (t) \cdot x$ and a variance about that mean is equal to $\sigma_x^2\ (t) = \sigma(t) \cdot x^2$. The simplest case where stationarity is assumed (i.e., $\lambda_x(t) = \lambda_x$, $\sigma_x^2\ (t) = \sigma^2 x^2$) yields a log-normal density function for the transition probabilities. Moreover with $\tau = 0$, and $x(o)$ the prescribed starting state the process has an expected trend of $M(t) = x(o)\ \exp\ (\lambda t)$ and variance $V(t) = (x(o))^2 \exp\ (2\ \lambda t)\ (\exp\ (\sigma^2 t) - 1)$. Note that the expected exponential trend is the same as the linear birth process model with constant birth rate λ, but the variances of the two processes are quite different.

Tintner and Sengupta (1972) give several examples of how the parameters can be made functions of one or more policy variables sometimes with lagged responses. They also show how the model can be extended to a N variable context where the instantaneous means of change and variances and covariances of change are of the form:

$$\lambda_i(t) = \lambda_i \cdot x_i(t) \qquad i = 1, 2, \ldots, N.$$

$$\sigma_{ij}^2\ (t) = \sigma_{ij}^2 \cdot x_i \cdot x_j$$

These assumptions yield a multivariate lognormal probability distribution with estimatible means, variances and covariances. Because of the simple assumptions the mean trends of each of the variables are expressions which do not involve values of the other variables. Interaction effects appear entirely in the covariance terms.

4.6. Urban Development as the General Epidemic Model

How does a town grow? Since in one sense the existence of towns is due to the external economies provided by individuals agglomerating in one place, development could have its basis here. But the very notion of external effects is a rather clumsy device to reconcile a theory of individual economic activity with the functioning group present in society. It is, as it were, the supplement which must be added to reconcile the summation of individual activity with the aggregate of the real world. Probably for this reason, externalities have only a loose connection with incisive logic of economic theory. If external economies be an artifact of an individualistic interpretation of the economy, presumably a theory better able to handle urban development in aggregate would treat the town as a functioning society with 'externalities' as integral parts of the functioning whole.

Locational decisions are based on views of the future. Cities grow because people are expecting them to grow and base their actions on this prediction. Individual expectations are formed on the basis of past performance and if past performance is poor, then a forecast of the future based on this is self-fulfilling. Yet how can individual firms forecast their development? Even in a fairly static urban economy, incessant change occurs with enormous rates of turnover of households in houses, of employees in a firm, of firms in an industry and of their shares in the total product. All this without any apparent change in the landscape. Thus in Ontario, the average birth rates of firms between 1961-65 was 7.0 and the death rate 6.1 with the machinery industry registering 11.8 and 6.6 and furniture 9.4 and 8.3 (Collins, 1970). Firms exhibit constant change in methods,

in products, in imports, in sizes; firms die; industries die. Consequently, the future is only poorly predictable by the individual entrepreneur. He may not even know what product he will be making five years from now. He does not know what skills he will have to draw on, what contacts he will require, what external support he will need. In such circumstances, the present environment sui generis may be of little significance and only in so far as it reflects future conditions will it be weighted. Consequently, location at two different places may be equally accommodating in terms of present needs but be quite different in terms of possible future requirements.

Even the firm with simple requirements for its present line of products and which could thus be footloose, will not usually believe that its products and requirements will be the same a few years from now. It may not be locationally tied in terms of current production but in order to maintain flexibility to change its line of goods, it should be near to a wide range of skills which might be used. In order to guarantee access to as wide a range of skills and products and markets which may be needed by him in the future, the entrepreneur will choose a location where these are presently available (i.e. are expected to be in the future). Thus even the 'footloose' firm may be quite demanding in terms of a present location geared to future unknowns. This line of reasoning has an interesting implication for the notoriously unreliable results of analyses of questionnaires on locational decisions. Clear cut answers to precise questions are very likely to give nonsensical results when the really truthful answer will be vague and look stupid.

It is not unreasonable to assert that the larger the town, the larger the variety of products and skills. This means a wider structure so that

new traits are more likely to be accepted because they are more likely to form a connection with the existing structure. This is particularly true of the manufacture of new products involving the latest technology because they will by definition involve specialist skills and knowledge found most readily in the larger towns. The unspecialized nature of the large city maintains its viability - in this sense, cities are not unlike biological species.

Labour migration is much more important than plant migration in the real world; this implies that productivity is a characteristic of place rather than plant techniques. However, invoking external economies to explain this is only tucking away our problems under a blanket phrase. In fact, using the terms in their commonly held meaning, plant techniques would seem to be of greater importance than external economies for very many firms. It would appear then that production functions are dependent on environment. But again the order of causation appears incorrect. It seems better to speak of the town or city as the unit, as a single cultural group, with both the exogenous environment and the endogenous techniques responding to the perceived future. Thus, both labour additions and technical changes occur together, they will covary in time and space.

To formalize these notions we shall make the immediate future the product of the present with change occuring from random disturbances. Forecasting then requires the fitting of a first-order autoregressive process and we are in a simple Markovian world. The reader is likely to be exasperated to find, after struggling through this long preamble, that the dynamics are as before. Nevertheless we do require two ideas.

(1) Decisions are made in terms of a predicted future: this allows the later reconciliation of planning models and positivistic models.

(2) The degree of uncertainty about the future is great and this favours large growing places.[2]

Let a worker entering the labour force of a city have the latest technical knowledge; as time passes he does not keep up to date, so that eventually his advanced skills are zero. He will attract new workers to enter the city because of his modern skills but there will be a point in time beyond which he does not attract at all. Assume that the period is an exponentially distributed random variable with mean μ^{-1}.

In any area having N persons there will be three classes of person at time T.

(1) $m(T)$ persons who do not inhabit the city

(2) $n(T)$ persons of the city who possess modern skills

(3) $N - m(T) - n(T)$ city dwellers who no longer possess these skills.

Let the system be in state (m,n) if $m(T) = m$ and $n(T) = n$.

$P_{m,n}(T)$ is the joint probability distribution of $m(T)$ and $n(T)$.

Let $P_{m,n}$ denote the probability that at some time there are m non-city dwellers and n 'spreaders'.

B - intensity of transmission between individuals

$\rho = \mu/B$ - relative removal rate; if large, forgetting is rapid

$$= \frac{\text{Average time for a randomly chosen pair to communicate}}{\text{Average length of time for which a spreader is active}}$$

If the process is considered only at those points of time when a change in state takes place, there is an imbedded random walk over the lattice points (m,n). From the state (m,n) there are two possible transitions:

---

[2]The following mathematical analysis has been adopted largely from Bartholomew (1967).

(1) $(m,n) \to (m-1, n+1)$, the attracting event,

(2) $(m,n) \to (m, n-1)$, the forgetting event,

except that $(m,0)$ is always absorbing. The random walk is Markovian because transition probabilities depend only on the present state of the system. The infinitesimal transition probabilities are:

(1) $Bmn\,\delta T$

(2) $\mu n\,\delta T$

and since they must sum to unity, we obtain for the associated probabilities

(1) $$\frac{Bmn\,\delta T}{\mu n\,\delta T + Bmn\,\delta t} = \frac{m}{m+\rho}$$

(2) $$\frac{\mu n\,\delta T}{Bmn\,\delta T + \mu n\,\delta T} = \frac{\rho}{m+\rho}$$

For the state (m,n) to have been reached, immediately previous states must have been $(m+1, n-1)$ or $(m, n+1)$. We may thus set up a difference equation

$$P_{m,n} = \left(\frac{m+1}{m+1+\rho}\right) P_{m+1,\, n-1} + \left(\frac{\rho}{m+\rho}\right) P_{m,\, n+1}$$

$$m \geq 0; \quad n > 1; \quad m + n < N + a$$

$$P_{m,0} = P_{m,1}\left(\frac{\rho}{m+\rho}\right) \quad ; \qquad 0 < m \leq N$$

$$P_{m,1} = P_{m,2}\left(\frac{\rho}{m+\rho}\right) \quad ; \qquad 0 < m \leq N$$

The initial condition is $P_{N,a} = 1$. For $m = N$ and $1 \leq n \leq a$, the probabilities are

$$P_{N,a-i} = \left(\frac{\rho}{N+\rho}\right)^i \quad ; \qquad i = 0, 1, \ldots, a$$

Those on the diagonal $m + n = N + a$ are given by

$$P_{N-i,\ a+i} = P_{N-i+1,\ a+i-1}\left(\frac{N - i + 1}{N - i + 1 + \rho}\right) \quad ; \quad i = 1, 2, \ldots, N$$

Numerical results may be obtained from these formulae. Figure 4.10 present the general features in terms of $n_H = N - m$, persons who inhabit the city and $a = 1$, the initial number of 'spreaders.' The terminal state depends critically on the relative sizes of N and ρ. With a threshold at $N/\rho = d = 1$. For $d \leq 1$, there is a high degree of positive skewness so that urbanization is unlikely to develop. For $d > 1$ there is a bimodal distribution, one near the origin and the other in the upper part of the range; variances, of course, are very large. As d is increased, the distribution becomes U-shaped and eventually is concentrated at N.

If we think of many areas, each with its initial and identical population and not competitive for labour, then any one of the graphs would depict the steady state probability that a person lives in a certain size of city, that a certain proportion of the regional population lives in a certain size of city.

While still very imperfect, this model does appear to capture some of the essence of urban development. In the first place, we compare the rate of communication of skills (conceived as a migration-education process) with the rate of technical change. An increase in a town's population can only come by being in the forefront of modern technology - this attracts workers. One has to attract new workers because these represent knowledge of the new technology when the initiators do not keep up to date. An urban population which has fallen behind technically cannot attract others. By concentrating on recent changes, it is clear that a forecast equation based on such a growth process would include the first few derivatives or differences. This is in line with practise. The importance of the stochastic element cannot

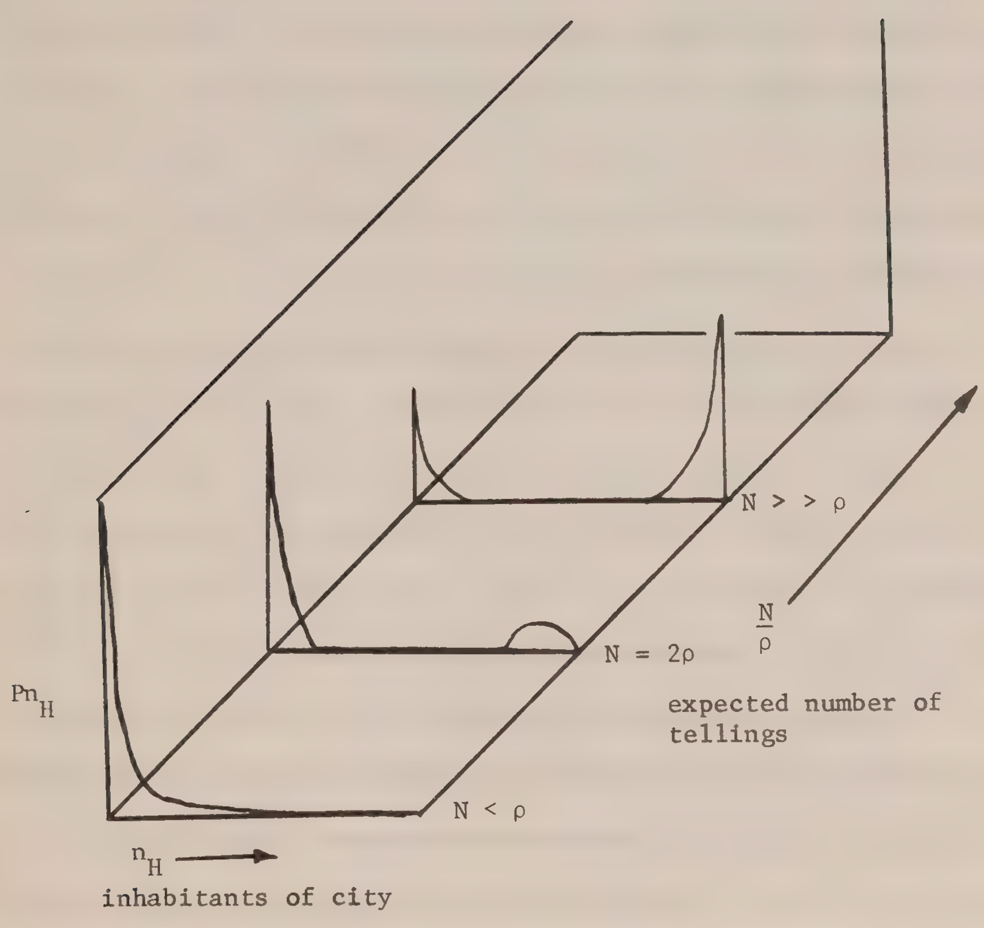

N > > ρ
N
ρ
N = 2ρ
Pn
H
expected number of
tellings
N < ρ
n
H
inhabitants of city

Figure 4.10

be overemphasized. While the given values of the various rates do define unique mean results as well are ranges of likely and unlikely results, there is still considerable variation in what can happen. This is particularly true where N/ρ is between 1 and 4 approximately.

However, there are other important aspects to urban development, largely concerned with space, which are not treated. As set out above, each city is independent. While clearly this cannot be true, the sort of dependence which should be introduced is not obvious. Completely conflicting postulates are equally defensible, in spite of the fact that this is a topic involving multi-million dollar policies.

To some extent cities must compete: certainly in the short run cities appear competitive for new developments, one city will absorb what might have gone elsewhere. Similarly labour going here cannot go there. Empirically in Ontario in decades when development is most rapid, growth is differentiated markedly at the local level, implying that some towns are gaining at the expense of their neighbours and perhaps using their labour. On the other hand, when development slows down, this local differentiation declines suggesting that there is local evening of growth. In the long run cities must provide mutual support rather than mutual exclusion: this may be saying no more than that they generate technical change for increasing productivity and that they support large consumer markets. Again empirically there is a strong "overspill" effect such that many towns of Southern Ontario grow basically because of their proximity to Toronto.

There is a strong tendency for government to assume that each potential development will occur somewhere and that location is a policy variable. The basis of this proposition as towns as competitors is not self-evident.

4.7. Deterministic Difference Equation Models of Urban Change: An Introduction

The mathematics of dynamic processes is generally written in differential equations, where time is continuous, and difference equations, where time is assumed to be discrete. A system of linear first order difference equations makes the claim that the current state of the system is a linear function of the state of the system in the previous time period. That is, for the system with one state variable $Y(t) = a_0 + a_1\ Y(t-1)$ where a_0 and a_1 are parameters. Thus, it is readily seen that the simple N-state Markov chain model is a system of N homogeneous difference equations:

$$\Pi_1(t) = P_{11}\ \Pi_1(t-1) + P_{21}\ \Pi_2(t-1) + \ldots + P_{N1}\ \Pi_N(t-1)$$

$$\vdots$$

$$\Pi_N(t) = P_{1N}\ \Pi_1(t-1) + P_{2N}\ \Pi_2(t-1) + \ldots + P_{NN}\ \Pi_N(t-1)$$

or in matrix form: $\Pi(t) = P\ \Pi(t-1)$.

Where $\Pi(t)$ is the vector of the state probabilities in the t^{th} stage of the process and P is the transition probability matrix. Similarly all of the other models of the previous sections have been derived using either difference or differential equations.

In this section, some simple deterministic difference equations are studied with a view toward appraising their potential utility for representing various urban processes. The purpose of such models is two-fold. First, they should give qualitative insights into dynamic mechanisms at work in cities. In order to achieve analytical solutions, it is usually necessary to make simplifying assumptions which deprive the models of a

strong empirical foundation. Thus, it is often the form of the solution which is important. The models are more in the nature of parables which can illustrate the consequences of fallacious reasoning. The second use of these models is of course to lay a stronger theoretical basis for subsequent, more realistic operational models.

In this and the following sections then, we attempt to characterize the basic strategy using of difference equation models and secondly to apply these models to urban problems suggesting the limitations as well as the benefits of such approaches. Although these models are not Markovian in the strictest sense in that they are deterministic rather than stochastic, their simple form of time dependency is certainly of a Markovian type.

The usual steps in difference equation modelling are: (1) identify the dynamics of the process in terms of lagged (and perhaps some concurrent) relationships between the variables; (2) reduce this system of equations to the minimal number of equations which contain all the information of the initial set (one equation in one unknown if possible); (3) determine the conditions for an equilibrium and determine the solution(s); (4) determine the general solution; (5) interpret these solutions in terms of the parameters.

The simplest form of difference equation, model is linear and first order. Oates, Howrey, and Baumol (1972) and Baumol (1972) use this model to study an urban problem. The problem, perhaps more relevant to U.S. cities than Canadian, concerns the flight of middle and high income households to the suburbs, a move which is both caused by and causes a physical deterioration in the urban housing stock. This cycle of cyclical causation continues until an equilibrium level of income and housing quality is

established. To adopt their notation, the basic model is:

$$Y(t+1) = r - sD(t) \qquad \text{with } s > 0$$

$$D(t) = u - vY(t) \qquad \text{with } u > 0$$

where $Y(t)$ is urban per capita income during time period t and $D(t)$ is some measure of deterioration during time period t. In reduced form, these equations can be represented by $Y(t+1) = r - su\ vY(t)$, a first order linear difference equation. The equilibrium level of income, if it exists is

$$Y_e = \frac{r-su}{1-sv}$$

Mathematically this equilibrium will be stable if $(sv)^2 < 1$ or more specific to this model if $0 < sv < 1$. The general solution, by which the time path to equilibrium may be traced is $Y(t) = (Y(0) - Y_e)\ (sv)^t + Y_e$. Thus, the equilibrium income is entirely a function of the parameters (the dynamics of the system) and is in no way affected by either the initial income or the level of deterioration. An effort to change the situation by altering incomes of housing quality will only be temporary. Ultimately, the incomes and housing quality will fall to Y_e and D_e respectively.

Now, one should be mindful of our earlier comments regarding the interpretation of equilibrium results. "Temporary" amelioration of the problem may be a quite effective way to deal with it depending on the rapidity with which the system progresses to the equilibrium (which in turn depends on sv) and the costs of changing the initial state and of course the costs and effectiveness of alternative remedies.

Regarding alternative remedies the authors suggest that the policy makers could interfere directly with the dynamics of the process, making for example $u = c_o - c_1E$ (expenditures) and $s = c_2 + c_3T$ (taxation).

Alternative government policies can be investigated to determine the sensitivity of equilibrium levels of income and deterioration to various combinations of taxes and expenditures. The authors point out that changes in T and E have multiplier effects on Y_e and D_e.

Now in both presentations of this model, the authors recognize the simplicity of their assumptions. In the second paper, Baumol (1972) formulates but does not solve a two region (city-suburb), two social group (rich and poor) extension. Many other variations on this basic model could be suggested. Two extensions will be sketched here.

First, note that the expenditures in the Oates, Nowrey and Baumol model are exogenously determined. An alternative approach would be to make these variables endogenous. One formulation could be:

$$Y(t) = r - sD(t-1)$$

$$D(t) = c_o - c_1 E(t) - vY(t)$$

$$E(t) = w_o + w_1 Y(t-1)$$

where E(t) is the level of expenditures, assumed to be linearly related to previous incomes; the reduced form equation is:

$$Y(t) = r - sc_o + sc_1 w_o + svY(t-1) + sc_1 w_1 Y(t-2)$$

a second order difference equation. A third order equation would result under the more realistic assumption that deterioration could be retarded by expenditures but with a one period lag. Making expenditures endogenous implies that the local government is severely restricted in its ability to stop the flight to the suburbs. In fact, departures to the suburbs accelerate deterioration both directly through income and indirectly through the decreased ability on the government's part to add in halting physical deterioration. Of course, the reverse case is equally true: if the initial income level is

less than Y_e, income will rise, reducing $D(t)$, and increasing $E(t)$ until Y_e, D_e, and E_e are reached.

Let us now turn to the solution of this model. The quilibrium solution is

$$Y_e = \frac{r - sc_o + sc_1 w_o}{1 - sv - sc_1 w_1}$$

One should note that the equilibrium may be altered in a variety of ways by changing one or more of the parameters. The most effective way to attain a desired equilibrium level, or perhaps some transient value, depends on the sensitivity of the results to changes in the parameters and the costs of such changes. In this connection, parameter w_1 can be interpreted as the tax rate. There are obviously political and economic limits to changing w_1 as well as the other parameters.

The second way to incorporate government expenditures endogenously within the model and yet retain more of an activitist stance, is to imbed the dynamics into an optimizing framework. (This discussion anticipates the subject of the next chapter, but it follows naturally from the presentation of the Baumol models.) To illustrate this modelling strategy, consider a small modification to the preceding model. Replace the third equation with:

$$E(t) = G(t) + w_1 Y(t-1)$$

It is now assumed then that expenditures are only partially dependent on income. $G(t)$ a government decision variable relating to deficit spending and revenues from other sources, has replaced the constant w_o. The reduced form dynamics of this system is:

$$Y(t) = r - sc_o + svY(t-1) + sc_1 w_1 Y(t-2) + sc_1 G(t-1)$$

The government may wish to determine the appropriate sequence of public expenditures G(t) to maximize income levels in year N or in years N-n, N-n+1, . . ., N. Of course the expenditures may be constrained in aggregate over the N year sequence and/or for each year individually. One set of constraints would refer to budgetary conditions; the other set would describe the dynamics of the process. In linear programming terms, the problem could be to maximize:

$$Z = \sum_{t=N-n}^{N} Y(t)$$

subject to the constraints:

$$\begin{aligned}
&Y(3) & -sc_1G(2) &= r-sc_o+sv\hat{Y}(2)+sc_1w_1\hat{Y} \\
&-svY(3) + Y(4) & -sc_1G(3) &= r-sc_o \\
&-sc_1w_1Y(3)-svY(4)+Y(5) & -sc_1G(4) &= r-sc_o \\
&-sc_1w_1Y(N-2)-svY(N-1)+Y(N) & -sc_1G(N-1) &= r-sc_o \\
&G(2)+G(3)+ \ldots +G(N-1) & &\leq B \\
&G(2) \;\ddots & &\leq B(2) \\
&\qquad G(N-1) & &\leq B(N-1)
\end{aligned}$$

$\hat{Y}(1)$ and $\hat{Y}(2)$ would be specified as the two most recent observations on the income variable. In addition, there are the usual non-negativity constraints: $G(t) \geq 0$.

Not only would this model yield an optimal solution for specific values of the parameters, it would also yield shadow prices indicating the sensitivity of Z to small changes in some linear combination of the parameters. For example, third dual variable would measure the increased payoff resulting from a unit

increase in the difference $r-sc_o$. The dual variables relating to the budgetary constraints are indicators of the amount Z would increase if one of the ceilings on expenditures were increased by one unit. A variety of post-optimal sensitivity analyses could be performed on such a problem. Other forms of the objective function could be used. Problems with non-linear functions could be solved using dynamic programming or optimal control methods. These methods are discussed in the following chapter.

There are very few additional examples of theoretical difference or differential equation models in the urban literature. Lachene (1965) postulates several alternative population growth accessibility relationships within a differential equation format, but does not carry the analysis very far. Hudson (1970) used the Lotka-Volterra predator-prey differential equations to look at a two region city-suburb migration process. Papageorgiou (unpublished manuscript at McMaster University) extends this to a multiregional case using the same equations with a migration being determined by a gravity model; this format quickly encounters computational problems making highly realistic models almost certainly intractable. The goal behind such analyses should be to elucidate certain urban issues in albut highly simplified contexts in order to gain qualitative insights into such processes and to aid in the ultimate construction of operational models.

Of course, not all deterministic difference equations are non-operational. In demography and ecology there are several models set up in matrix form which closely resemble the finite Markov chain model except that the matrices are not stochastic. (See Usher (1972), Keyfitz (1968)

and Rogers (1968) for reviews and examples of such models.) A matrix operator A is a square matrix of a very simple structure. It accomplishes two things when multiplied by the population age structure matrix: it ages the existing population and, by using birth and death rates for different age groups, it causes individuals to be added and subtracted from the population. The basic model is thus:

$$a_{t+1} = A \cdot a_t$$

where a_t is a vector indicating the number of individuals in each age group at time t

an A is a matrix the first row of which is composed of fertility rates f_j for each age group and each of the other rows contains one non-zero element $P_{i,\ i-1}$ which is the proportion of individuals in the $i - 1$ age class that will survive to the i^{th} class in one time period, the time periods being chosen in such a way that all others must necessarily perish.

This class of model, proposed by Leslie (1945), has had many ecological applications and has been extended to population growth and interregional migration. In the migration case, A is simply expanded to include in and out migration as well as births, deaths, and aging. Rogers (1968) and Stone (1968) have applied such models with some success in describing growth and migration in California and Canada respectively. These models in their simplest forms are highly empirical and many of the same criticisms applied to empirical Markov models concerning their lack of sensitivity to theory and policy can be made of these approaches. These objections can be answered in much the same way as well.

Of particular interest in this regard are some recent developments in ecological models which attempt to determine values of the parameters taking into account certain additional relationships (Usher, 1972). The first determines values for a subset of parameters by making them a function of the principal eigenvalue (equilibrium growth rate) of the system. The eigenvalue and some of the parameters are mutually determined; thus iterative methods are used to solve for both. Although the nature of the problem (forest harvesting) would seem to limit any direct analogies, it might be worthwhile to attempt to formulate urban models which simultaneously determine overall equilibrium growth rates and parameter values.

Of more obvious relevance are those models which allow their parameters to vary through time in a way which is determined by the current state of the system. This essentially extends the simple birth-death and queuing theory state dependent parameter models discussed earlier in this chapter, where rate of change is influenced by economies of scale and eventually perhaps crowding. In the ecological example, fecundity is assumed to fall off exponentially with total population size while the survival function has an inverted sigmoid shape. Although it is unlikely that urban systems would have functions identical shape, it is useful to note that it is a relatively easy matter to deal with such non-stationary processes. The results, of course, are not as tidy as stationary models. No general or closed form expression of transient or equilibrium behaviour is available, but the expected trajectory of alternative initial conditions (population size and age structure in the simplest case) and different fecundity and survival functions can be determined and plotted. Depending

on the selection of initial conditions and parameter functions, sigmoid or logistic growth curves with dampened or regular oscillations may be generated (see Usher, 1972). In the urban case birth rates could be made to decline and in-migration rates to increase with increasing city size. Having estimated these parametric functions statistically or perhaps speculatively, the spatial change within an urban system could be simulated.

4.8. Economic Models of Urban Growth: Difference Equation Approaches

There are several concepts by which current models have been classified. In this section, a classificatory distinction is to be made between growth models which may be termed 'factor-recursive' from those defined to be 'factor-passive.' The term 'factor' is meant to include both capital and labour (which are spatially mobile to some degree) but exclude land. A factor-recursive model may be defined as one which has (a) well-defined hypotheses about the inflow or outflow of factors into or from an urban factor market and (b) well-defined hypotheses about the effects of factor shortages or surpluses on the growth of population or income. A factor-passive model is any other.

The principal advantage of using the factor-recursive - factor-passive classification is that it permits us to put a particular emphasis on the question of the cause of growth. The general hypothesis to be developed below is that there likely exists no valid concept of optimum urban size within a static framework. Urban growth is seen to occur at least partly in response to factor surpluses or shortages. It will be argued that there are strong but not complete inter-relationships between factor markets and urban growth that make urban growth and its dynamics fundamentally dependent on the dynamics of factor markets.

This growth hypothesis is not without precedent in the literature. Chinitz (1971), for one, has formulated a similar hypothesis in slightly different terms. He asserts that national economic growth does not wholly determine local urban growth as is often assumed in empirical urban and regional growth models. Local factor surpluses or shortages cause local

growth to be at least partly independent of national trends. Further, Chinitz asks whether an area grows because of inherent locational advantages or because past growth enables it to develop certain assets? In other words, is it strictly true that local competitiveness unidirectionally leads to economic growth or does economic growth significantly affect competitiveness?

4.8.1 Factor Passive Models

Initial factor-passive models of urban growth were of two varieties. One is the export-base model which emphasizes the interrelationships between population and employment. The other is the Keynesian income model which emphasizes linkages among expenditure categories. More recent models have attempted to integrate employment, population, and expenditure concepts.

Export-base models are built on the assumption that urban growth is determined by the rate of export-sector growth. Czamanski (1964) presents a typical static version of this model. Letting current population, total employment, complementary employment, export-sector employment, and urban-oriented (or local) employment be denoted by P, E, E_c, E_e, and E_u respectively, Czamanski expresses the model as follows.

(1.a) $$P = a_1 + b_1 E \qquad b_1 > 0$$

(1.b) $$E = E_e + E_c + E_u$$

(1.c) $$E_c = a_2 + b_2 E_e \qquad b_2 > 0$$

(1.d) $$E_u = a_3 + b_3 P \qquad 1 > b_3 > 0$$

This model can be reduced to a single equation relating population to export employment.

(2) $$P = \frac{a_1 + b_1 a_2 + b_1 a_3}{1 - b_1 b_3} + \frac{b_1(1+b_2)}{1 - b_1 b_3} E_e \qquad \text{assuming } b_1 b_3 < 1$$

The deficiencies of this simple export-base model are discussed in Appendix B. Here we may summarize these deficiencies by saying that the model is too rigid and simple. It is a purely demand-oriented model which takes no account of factor supplies. Migration of capital and labour in and out of the urban centre must occur simultaneously to satisfy the conditions (1.a) to (1.d).

In addition the model begs the question of what causes growth in export employment. While it may well be that the export sector responds at least in part to conditions external to the individual city, it seems reasonable to expect that the size of the export-sector is responsive to the size and other attributes of that community. To the extent that such a two-way linkage exists between city size and export employment, the simple export-base model fails to be a valid growth model.

Some variations on a dynamic form for the export-base model have been attempted. Czamanski (1965), among the first to develop such a model, had the limited aim of allowing for time lags in each of the relationships in (1.a) through (1.d). Using data for the Baltimore SMSA, he identified the best lag values in terms of statistical goodness-of-fit. There are no theoretical justifications given for his choices but some reasonable supporting arguments might be made for each. Letting the subscript 't' refer to a time period, his dynamic version is as follows.

(3.a) $$P_t = a_1 + b_1 E_{t-2} \qquad b_1 > 0$$

(3.b) $$E_t = E_{e,t} + E_{c,t} + E_{u,t}$$

(3.c) $$E_{c,t} = a_2 + b_2 E_{e,t} \qquad b_2 > 0$$

(3.d) $$E_{u,t} = a_3 + b_3 P_{t-1} \qquad 1 > b_3 > 0$$

This model can be solved as a difference quation relating population to export sector employment.

(4.a) $$P_t = c_0 + c_1 E_{e,t-2} + c_2 P_{t-3}$$

where

(4.b) $$c_0 = a_1 + b_1 a_2 + b_1 a_3 \qquad c_1 = b_1(1+b_2) \qquad c_2 = b_1 b_3$$

Czamanski's dynamic model produces a temporal sequence of events which is consistent with empirical observation. A maintained increase in the export-sector employment leads to an increase in population after some period in time. The increase in population subsequently leads to an increase in urban-oriented employment and to still-further indirect increases in population and urban-oriented employment.

The dynamic export-base model thus permits some flexibility in a temporal sense. The instantaneous adjustment of population and employment has been muted. However, the dynamic version still takes no direct account of factor supplies and does not answer the question of the cause of export growth.

Keynesian models of urban growth with their emphasis on incomes and expenditures are underdeveloped in comparison to export-base models. In part this has been due to the interest of regional planners specifically in population and employment growth. Keynesian models have begun to create interest however with the development of dynamic versions capable of describing inter-urban transmission of business cycles.

We may begin by defining a simple static income model which ignores the government sector. Letting Y, C, I, and E denote urban output, consumption, investment, and net exports respectively, the expenditure identity and multiplier hypothesis may be expressed as follows.

(5.a) $$Y = C + I + E$$

(5.b) $$C = a_0 + a_1 Y$$

By substitution, the familiar multiplier model is derived.

(5.c) $$Y = \frac{a_0}{1-a_1} + \frac{1}{1-a_1} (I + E)$$

The deficiencies of this model are similar to those for the export-base model. It is too rigid and simple and fails to identify other interrelationships among Y, I, and E.

A dynamic Keynesian model helps to reduce the rigidity of the static model as in the case of the export-base model. Letting the subscript 't' again denote a time period, one common form of the multiplier accelerator model can be expressed as follows.

(6.a) $$Y_t = C_t + I_t + E_t$$

(6.b) $$C_t = a_0 + a_1 Y_{t-1}$$

(6.c) $$I_t = b_0 + b_1 (Y_{t-1} - Y_{t-2})$$

This model can be reduced to a second-order difference equation in Y_t.

(6.d) $$Y_t - (a_1 + b_1)\, Y_{t-1} + b_1 Y_{t-2} = (a_0 + b_0) + X_t$$

Casetti (1969) has used a model similar to this to predict the effects of net export fluctuations on regional income. Depending on the values of the parameters in (6.d), a once-and-for-all change in X may induce a cyclical temporal trend in Y. More generally, any temporal pattern in exports may be replicated, augmented, dampened, or otherwise distorted as a temporal pattern in urban income depending only on the regional parameters. In this manner, national or regional business

cycles, 'imported' into the city as export fluctuations may result in a distinct pattern of city income and output variations over time.

Jutila (1973) presents an alternative to the model (6.a) through (6.c) in which the variables are dynamic through time and space. In this model, investment is determined not only by temporal relationships as in (6.c) but also by the spatial position of the city within a region or within a hierarchy of cities. Spatial position enters Jutila's model as an explicit attempt to represent the spatial diffusion associated with innovation and economic growth. He fails, however, to provide an explanation for why growth is diffused spatially according to any particular pattern.

Initially, Keynes formed his model to explain national macroeconomic conditions at a time of high unemployment combined with wage-price inflexibility. As a short-run model of the city with high unemployment and low factor mobility, the Keynesian model appears to be appropriate. It fails however as a longer-run model when labour and capital can migrate because it does not allow for labour shortages. This is as true of the dynamic as of the static version.

The principal advantage of the dynamic Keynesian model over the static version is that capital growth or migration is specifically allowed for in the investment function. This represents a first step toward factor-recursive models of urban growth and is a step beyond the export-base model.

Several attempts towards a fusion of export-base and income models into a single model of population, employment, and income have been made. A number of researchers have estimated such models for urbanized regions and we choose to examine Bell's (1967) model of the Massachusetts economy as an example.

Gross National Product serves to determine regional income and output

through direct effects on export sales and indirect effects on the production of local services and capital exports. Regional output, together with a time trend and lagged capital stock, act to fix the size of the current capital stock. The wage rate, related solely to a time trend, and the current capital stock then establish the demand for labour. The demand for labour and the level of employment are equated so that the model implicitly assumes no shortages in the labour supply. The labour supply is determined by demographic changes in the lagged population together with migration which responds to a projected rate of unemployment. A schematic of this causal structure is presented in Figure 4.11.

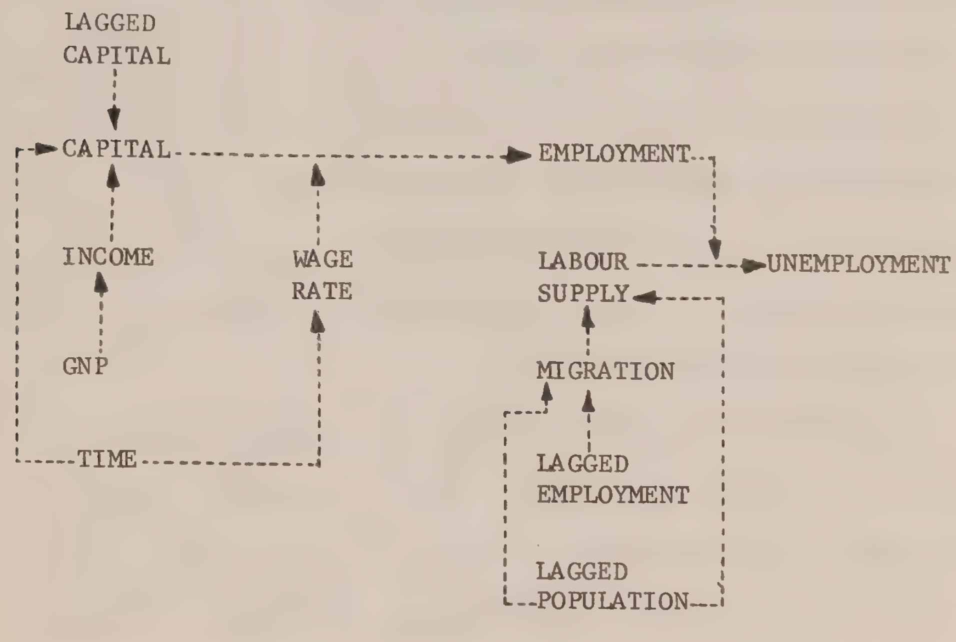


Figure 4.11 Causal Structure in Bell's Massachusetts model.

Note: All variables current unless otherwise noted.

Source: See text.

There are well-defined hypotheses about factor movements in and out of the region in Bell's model. Both capital and labour are hypothesized to have lagged-adjustment responses to changes in factor demand. Letting K_t denote the capital stock and Q_t denote regional output in year 't', the functional equation explaining capital stock takes the following form.

$$(7.a) \qquad K_t/K_{t-1} = f(Q_t, K_{t-1}/K_{t-2})$$

The migration component of labour force has a similar adjustment form. Letting M_t denote migration, E_t denote employment, and $L_{e,t}$ denote expected labour force, the migration hypothesis is given by

$$(7.b) \qquad M_{t+1} = a_0 + a_1(E_t - L_{e,t}) \qquad a_1 > 0$$

Thus, both capital and labour are seen to move in and out of the region in a way which ensures that factor supply and demand will tend to equalize over time.

There are not, however, well-defined hypotheses about the effects of factor surpluses or shortages on the urban economy for both capital and labour. More particularly, capital shortages and surpluses do have an hypothesized effect on the economy if these misalignments are measured in a special way but labour misalignments do not.

Commodity shortages or surpluses are usually explained in terms of price inflexibility. Within Bell's model, there is neither an explicit price for capital nor an explicit assumption about price flexibility. There is, however, some notion of an equilibrium rate of growth of capital in equation (7.a) and actual deviations from this growth rate may be termed shortages or surpluses in this sense. Such capital misalignments lead to different levels of employment than would occur in their absence. Thus, capital surpluses and shortages do have an effect on the economy generally.

There is, alternatively, a well-defined price for labour in the model. This price is totally inflexible within the model so that we can think of shortages and surpluses in the usual sense. However, labour surpluses or shortages in one period merely affect migration flows. There are no repercussions from labour misalignments back on the capital stock, incomes, outputs, or wages. In effect, the model merely serves to gradually reallocate labour surpluses out of the region without allowing for any effect on the local economy.

4.8.2. Semi Factor-Recursive Models

The notion of factor-recursiveness is relatively new in mathematical urban growth models. Most recursive models are still in a preliminary, abstract, and theoretical form.[3] Further, most of these models are recursive only in terms of one factor, usually labour, while the other is ignored. Those factor-recursive models which ignore one of the two factors are here defined to be semi factor-recursive.

Ghali (1973) presents a simple recursive model of income and capital growth which he empirically applies to the state of Hawaii. Letting Y, X, I, E, D, and Z be gross regional product, exports, investment, employment, bank deposits, and U.S. GNP respectively and letting corresponding uncapitalized letters represent annual relative rates of change, the basics o Ghali's model may be represented as follows:[4]

---

[3] Empirical applications, especially of models recursive in labour, are rare. Muth (1971) is a notable exception.

[4] Ghali's original presentation includes an additional exogenous variable in (8.c) which is immaterial to the present discussion.

$$(8.a)\qquad y_t = a_{10} + a_{11}i_t(I/Y)_{t-1} + a_{12}x_t(X/Y)_{t-1} \qquad a_{11}, a_{12} > 0$$

$$(8.b)\qquad x_t = a_{20} + a_{21}z_{t-1} \qquad a_{21} > 0$$

$$(8.c)\qquad i_t = a_{30} + a_{31}d_{t-1} \qquad a_{31} > 0$$

$$(8.d)\qquad D_t = a_{40} + a_{41}Y_t \qquad a_{41} > 0$$

$$(8.e)\qquad e_t = a_{50} + a_{51}y_t \qquad a_{51} > 0$$

This model is recursive in capital. The growth rate of bank deposits affects the growth rate of investment which in turn affects the growth rates of capital, income, and employment. A connection has been made between the accumulation of local savings and the act of local investment even though there is no explicit notion of capital surpluses or shortages in the model.

What Ghali suggests is that income and capital growth can be generated independently of the level of exports. This is not a new concept for an economic growth model, but it is quite different from the economic base hypothesis.

Ghali's model is deficient in at least two important respects. First, there is no explicit account taken of labour surpluses or shortages in this model. Employment growth is determined in (8.a) solely by the rate of growth of income with no repercussions on the remainder of the system. Secondly, Ghali provides no theory as to why capital, in his model , is regionally immobile. He does not answer the question of why local savings are re-invested locally when extra-regional investment may be more profitable.

Several current semi factor-recursive urban growth models are descendents of the simple export-base hypothesis. Niedercorn (1963) was among the first to develop such a model. Letting basic (export) employment and population in year 't' be represented by $E_{b,t}$ and P_t respectively, his

model asserts the following.

(9.a) $R_t = a_{11}R^*_{t-1} + a_{12}R_{t-1}$ $\quad a_{11}, a_{12} > 0$

(9.b) $S_t = a_{21}R_t + a_{22}$ $\quad a_{21}, a_{22} > 0$

(9.c) $R_t = (E_{b,t} - E_{b,t-1})/E_{b,t-1}$

(9.d) $S_t = (P_t - P_{t-1})/P_{t-1}$

(9.e) $R^*_t = (E^*_{b,t} - E_{b,t})/E_{b,t}$

(9.f) $E^*_{b,t} = a_{31}P_t$ $\quad 1 < a_{31} < 0$

The current growth rate of basic employment, R_t, is determined in (9.a) by the lagged growth rate and by the relative divergence, R^*_{t-1}, of the actual basic employment from some 'desired level', $E^*_{b,t}$. The population growth rate, S_t, is linearly related in (9.b) to the basic employment growth rate.

This model is, in some sense, factor-recursive in labour. There exists some 'desired' level of employment, related to population size in (9.f), which might be thought of as roughly the basic sector's potential labour supply.[5] In this sense, 'shortages' or 'surpluses' of labour are possible and these are reflected as negative and positive values for R^*_t respectively. Thus, labour shortages (or surpluses) so defined have a negative (or positive) effect on employment growth rates in (9.a) and a corresponding indirect effect on subsequent population growth in (9.b). There is a well-defined hypothesis about the effects of labour shortages or surpluses on the growth of regional population and employment. Further,

---

[5] The term 'potential' labour supply is used here because there is no reference to the local or service sector. Without knowing the employment characteristics of this sector, it is difficult to speak of an actual basic or local sector labour supply.

(9.b) represents an hypothesis about the movement of labour in and out of the region in response to labour market conditions. These two conditions establish that Niedercorn's model is semi factor-recursive under our definitions.

The anti-equilibrium nature of Niedercorn's model is of interest. The dynamic properties of his model have been investigated elsewhere and it may be concluded from those sources that this model possesses neither a static equilibrium solution nor a balanced growth (or dynamic equilibrium) solution. A verbal explanation is that, without any limits on population change either through natural increase or migration, there is a built-in multiplier in (9.a) and (9.b) which drives population and export employment to infinity with an increasing divergence.

The population growth equation (9.b) is at the source of this disequilibrium. In particular, the responsiveness of population growth to employment growth, the accelerator mechanism, is the key to the process. If a_{31} is zero, the population growth rate is exogenously given and is independent of the employment growth rate. The larger a_{31} is relative to a_{32} the more dependent is population growth on employment growth.

Niedercorn's model emphasizes the concept that the dynamics of the labour market are the key to understanding the dynamics of urban growth. If export-sector growth is not tied to external conditions and if any level of employment can be supported, then understanding migration behaviour for labour and capital is central to understanding how growth occurs. If migration patterns under these conditions are tied strongly to past employment growth, then there can be no concept of static equilibrium or balanced growth which is necessarily applicable. Stability in this case depends

critically on having migration behaviour which is to some significant extent independent of employment growth.

Paelinck (1970) has formulated a model which is similar to that of Niedercorn. Letting E, L, and P represent employment, labour force, and population respectively, his model takes the following form.

$$(10.a) \qquad E_{t+1} - E_t = a_{10}E_t + a_{11}(L_t - E_t) + a_{12}P_t \qquad a_{10}, a_{11}, a_{12} > 0$$

$$(10.b) \qquad L_{t+1} - L_t = a_{20}(P_{t+1} - P_t) + a_{21}(E_t - L_t) \qquad a_{21}, a_{20} > 0$$

$$(10.c) \qquad P_{t+1} - P_t = a_{30}E_t + a_{31}P_t \qquad a_{30}, a_{31} > 0$$

Employment increases in (10.a) in response to the sizes of population and employment and to the labour market surplus $(L_t - E_t)$.[6] The labour force increases in (10.b) with increases in population and it decreases with an increase in the labour market surplus. Population growth is related to the absolute size of employment and population in (10.c).

This model is semi-factor-recursive in labour. Labour is hypothesized to move in and out of the labour market in response to shortages and surpluses in the labour market. Further, the demand for labour is hypothesized to adjust to labour market misalignments in (10.a).

There is a difference between the Niedercorn and Paelinck models in terms of how labour supply adjustments occur. In Paelinck's model, population growth does not respond directly to labour market misalignments. Direct adjustment to such misalignments is taken up in a variable labour force participation rate. Further, this model does not allow for any

---

[6] Again, there is an implicit assumption of wage rigidity which permits labour force and employment to diverge.

decline in population.[7] In Niedercorn's model, however, population growth responds directly to employment growth which, in turn, is tied to labour market misalignments. This implies that labour surpluses or shortages are partly taken up by labour migration out of or into the region and allows for the possibility that population may decline. Niedercorn's model contains an accelerator mechanism through which population growth responds to labour misalignments in contrast to Paelinck's model where such a mechanism is considerably muted by the use of a variable participation rate.

4.8.3. A Simple Model of Urban Labour Market Dynamics

Paelinck, as does Niedercorn, suggests that what determines the dynamics of urban growth are the dynamics of the labour market. Their models are reasonably complicated however and it is useful to consider some simpler models of labour market dynamics both to better understand the implicit assumptions and to examine the dynamic nature of the solution. To this end, a number of exploratory models are now considered.

Begin by imagining that an urban labour market consists, in any time period, of three decisions and three decision-making groups; (i) the hiring decision made by a group of entrepreneurs, (ii) a job-seeking decision made by a group of job-seekers, and (iii) a population decision also made by the group of job-seekers. The initial model has three equations; one representing the decision of each group.

One equation asserts that entrepreneurs, as a group, decide how many jobs to offer on the basis of the expected population in that period. It

---

[7] Since a_{30} and a_{31} are positive in (10.c).

is assumed that, while this decision is being made, entrepreneurs have an infinitely elastic demand for labour at a given wage. However, once this hiring decision has been made, at the beginning of the current period, the entrepreneurs are assumed to draw up a production schedule. This schedule defines a sequencing of production activities which cannot be varied, or more especially increased, over the current time period and this renders the demand for labour totally inflexible at the predetermined level. In this way, the demand for labour can be seen to be suitably elastic from one time period to the next while being inflexible during any one period.

A second equation represents the hypothesis that the job-seekers, as a group, respond to the expected number of job-offerings in a period. The labour force is quite fluid by assumption. Job-seekers move in and out of the urban labour market over successive time periods in a search for employment. It must further be assumed that the urban centre's growth is uncorrelated with the growth of other urban centres nearby so that a job-seeker who fails to find a job in one centre will be motivated to seek employment elsewhere in another period.

The final equation asserts that the size of population is related to the number of job-seekers.

At a broad scale, there are at least two major difficulties with this model which are not resolved in this study. First, it is not clear what the implications are of assuming that entrepreneurs and job-seekers act as cohesive groups when observably they do not. Secondly, by connecting a city's labour force and population to its employment level, we exclude among other things the possibility that the city may be one part of a spatially-compact urbanized region with extensive inter-city commuting. The second difficulty merely restricts the range of application of the

model while the first, a difficulty implicitly shared by Paelinck and Niedercorn, is more serious.

This model may be specified in a more particular fashion. Letting P, E, and L be the population, job offerings, and job-seekers respectively, one form for this model is as follows.

(11.a) $\quad E_t = a_{10} + a_{11} g_1 P_{t-1} \qquad g_1 > 0 \qquad 1 < a_{11} < 0$

(11.b) $\quad L_t = g_2 E_{t-1} \qquad g_2 > 0$

(11.c) $\quad P_t = a_{20} + a_{21} L_t$

(11.d) $\quad a_{21} = 1/a_{11} \qquad a_{20} = -a_{10}/a_{11}$

The constant g_1 (or g_2) represents the growth rate for population (or job-offerings) expected by the entrepreneurs (or job-seekers). The relationship between P_t and L_t is assumed to be fixed over time and known by both entrepreneurs and job-seekers and this accounts for (11.d).

In an analysis of dynamic models, two main properties are often emphasized; convergence and balanced growth. A convergent dynamic model possesses endogenous variables which tend to converge to some fixed equilibrium value over time. A balanced-growth model possesses endogenous variables which, in a dynamic equilibrium, tend to grow at the same fixed rate. Observation suggests that convergent urban growth (or decline) is not common and that there appears to be no static equilibrium relationships among the growth rates of population, employment, and labour force. New properties have to be defined which are appropriate to the evaluation of urban growth models.

Here, we choose to emphasize two different properties of 'well-behaved' urban growth models; consistency and stability. Stability is used here to

denote a model whose endogenous variables do not display explosive temporal oscillations when exogenous variables are fixed at arbitrary levels. Consistency denotes a condition in which those parameters in a model representing expectations about growth rates are realized.

The present model is generally consistent but is unstable when growth occurs. Both dynamic properties depend on the geometric average of the expected growth rates, i.e. $\sqrt{g_1 g_2}$. If this average is less than unity (as would occur when both job-seekers and entrepreneurs expect a decline), population, labour force, and job offerings tend to decline over time with dampened cycles implying stability in decline. If $\sqrt{g_1 g_2}$ is greater than unity, all endogenous variables increase over time but with anti-dampened cycles. Growth results in instability. Consistency in either a growth or decline situation always occurs when g_1 and g_2 are equal regardless of the absolute growth expectation level. In other words, a sufficient condition for a certain rate of growth to occur in the model is that both entrepreneurs and job-seekers expect such a rate.

That it is impossible to have consistent growth without instability suggests a deficiency in this model. Admittedly, the model is a gross simplification of the growth process. Is there, however, a single element whose inclusion in the model would create a dynamic solution more consistent with casual observation?

At least three possibilities seem evident. First, anti-dampened cycles may be occurring because the labour force participation rate is not varying as in Paelinck's model. This can be tested by modifying the population-labour force relationship in (11.c). Secondly, the presence of anti-dampened cycles indicates ever-increasing migration in and out of the community. It is reasonable to expect that migration will, in part,

be sensitive to housing conditions. Suppose that a housing sector is introduced into the model and that the job-seeking decision is tied to housing conditions as well as to job forecasts. Stability may be introduced into the model due to the longevity of the housing stock. Thirdly, anti-dampened cycles may arise because entrepreneurs and job-seekers are forecasting by simple extrapolation when they could be expected to pay closer attention to the most recent past changes in job offerings and labour supply. The effect of this can be tested by replacing (11.b) for instance by a job-seeking decision which relates labour supply to an adaptive extrapolation of the change in population over two previous periods.

4.8.4. Variable Participation Rates and Urban Growth

Considering the first of the above suggestions, a possible redefinition of the initial model, (11.a) to (11.d), is as follows.

(12.a) $E_t = a_{10} + a_{11}g_1P_{t-1} \qquad 0 < a_{11} < 1 \qquad g_1 > 0$

(12.b) $L_t = g_2E_{t-1} \qquad g_2 > 0$

(12.c) $P_t = a_{20} + a_{21}L_{t-1} + a_{22}(L_t - L_{t-1})$

(12.d) $a_{20} = -a_{10}/a_{11} \qquad a_{21} = 1/a_{11} \qquad a_{22} < a_{21}$

In this model, a distinction is made between net new job-seekers $(L_t - L_{t-1})$ and old job-seekers (L_{t-1}). If the number of job-seekers increases (or decreases) in a period, it is hypothesized that the new net inmigrants to (or outmigrants from) the urban labour market carry with them a smaller number, a_{22}, of marginal dependents per job-seeker than do the established job-seekers. This has the effect of raising or lowering the average and marginal participation rates depending on the rate of growth of the labour supply.

The dynamic properties of the model depend on the size of a_{22} relative to a_{11} and on the growth expectations g_1 and g_2. Unstable areas of the g_1g_2 - $a_{22}a_{11}$ space are displayed in Figure 4.12. There are broad ranges of values for a_{22} which result in stability at commonly observed growth rates.

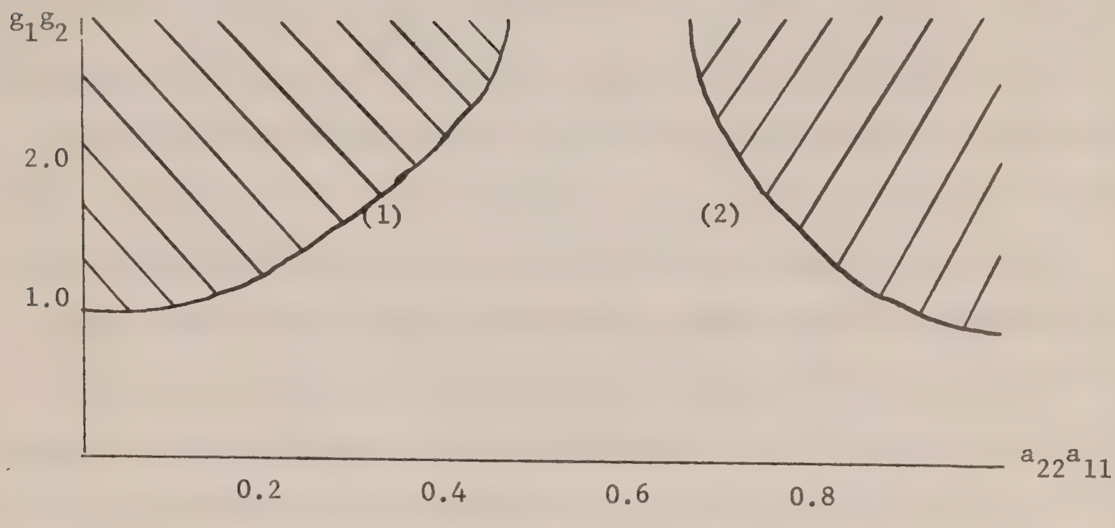


Figure 4.12 Stable region for model (12.a) to (12.d)

Notes: Hashed-in areas are unstable (explosive oscillatory).

(1) $Y = (X + \sqrt{X^2 + 4(1-X)^2})/2(1-X)^2$

(2) $Y = 1/(2X-1)$

where $Y = g_1g_2$, $X = a_{22}a_{11}$

Consistency is difficult to evaluate in this model. There is one consistency path along $a_{22}a_{11} = 1$ when g_1 and g_2 are equal but this implies that consistent decline ($g_1 = g_2 < 1$) is stable but consistent growth ($g_1 = g_2 > 1$) is unstable. This is not surprising since the assumption that $a_{22}a_{11} = 1$ reduces the model (12.a) to (12.d) to the model (11.a) to (11.d).

There may or may not be other consistency paths in $g_1g_2 - a_{22}a_{11}$ space but we have not been able to establish this either way as yet. There is an intuitive reason however for suspecting that there may not generally be consistent solutions when $a_{22}a_{11} < 1$. In (12.a), entrepreneurs make forecasts of future job-seekers based on a constant marginal participation rate (a_{11}). The further g_1 and g_2 are from unity, the further will be the actual marginal participation rate from a_{11}. There thus appears to be no nonzero expected growth rates (g_1, $g_2 \neq 1$) which sustain an equivalent realized growth rate when $a_{22}a_{11} < 1$.

This model suggests that the variable participation rate hypothesis may not possess suitable dynamic properties. While the hypothesis does provide broad parameter regions within which stability is present, there does not seem to be much room for consistency in conjunction with stability. This is especially true in the case of positive expected growth rates.

4.8.5. The Housing Sector and Urban Growth

Taking the second suggestion in section (d) above, we can as an example redefine the initial model, (11.a) to (11.d), as follows.

(13.a) $$E_t = a_{10} + a_{11}g_1P_{t-1} \qquad 0 < a_{11} < 1 \qquad g_1 > 0$$

(13.b) $$L_t = (1-p)g_2E_{t-1} + pa_{22}(a_{21}H_{t-1} + a_{20}) \qquad 0 < p < 1 \qquad g_2 > 0$$

(13.c) $$P_t = a_{30} + a_{31}L_t$$

(13.d) $$H_t = a_{41}H_{t-1} + a_{42}P_t$$

(13.e) $$a_{30} = - a_{10}/a_{11} \qquad a_{31} = 1/a_{11}$$

(13.f) $$a_{20} = a_{10}/a_{11} \qquad a_{21} = (1-a_{41})/a_{42} \qquad a_{22} = a_{11}$$

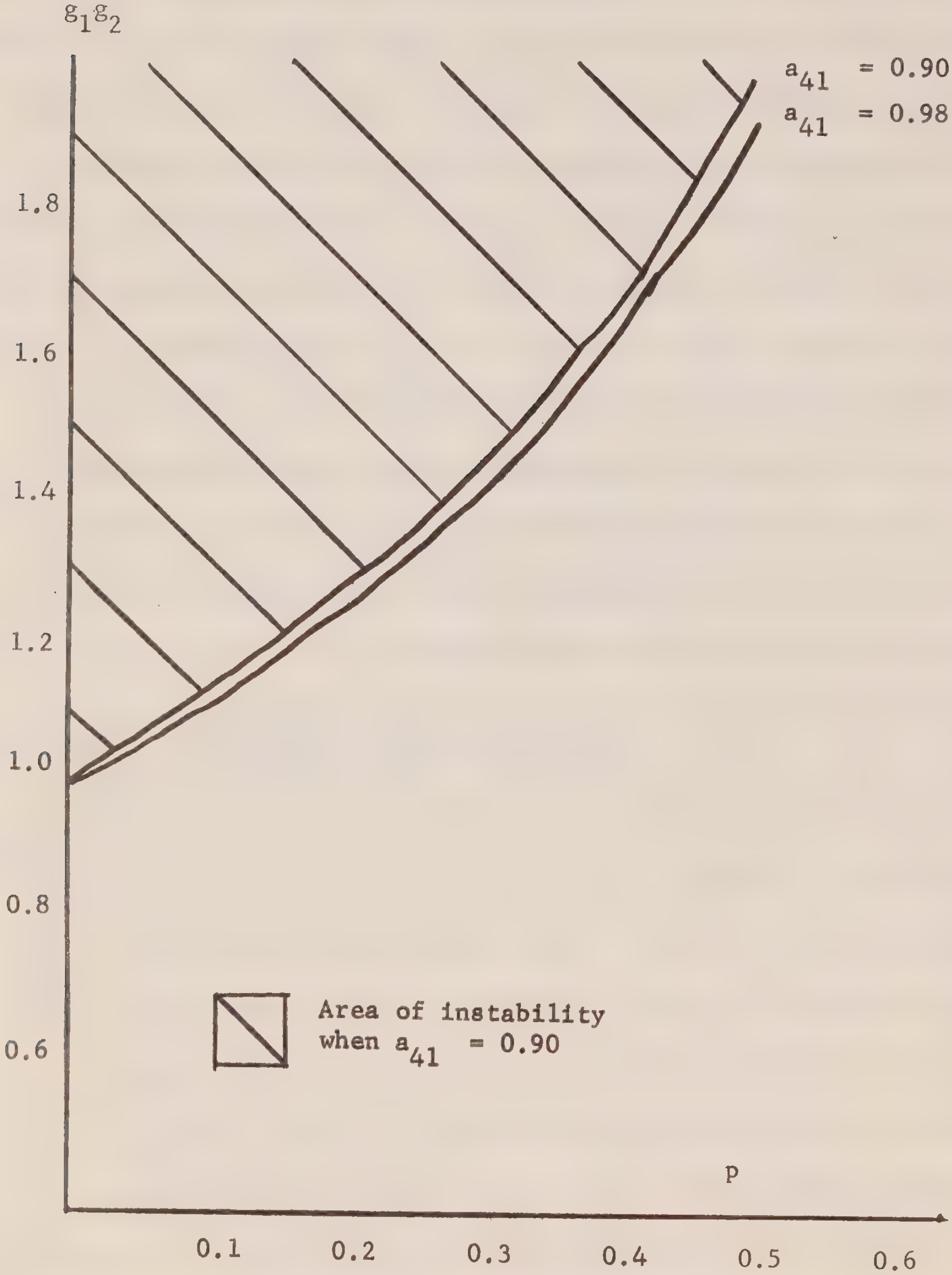
Job-seekers respond, in (13.b), to the expected number of job offerings,

g_2E_{t-1}, and to the housing stock. Defining $(1-a_{41})$ to be the rate of depreciation of the housing stock, it is seen from (13.d) that, for a temporally-fixed population, P, the housing stock eventually moves to an equilibrium value, H, such that $P = a_{21}H$. Thus, $a_{21}H_{t-1}$ in (13.b) is the population associated with a housing stock of H_{t-1} in equilibrium. The quantity $a_{22}(a_{21}H_{t-1} + a_{20})$ is thus the labour force associated with a housing stock of H_{t-1} in equilibrium. Therefore, it is hypothesized in (13.b) that job-seekers respond to a weighted average of anticipated job offerings and the labour force supportable by H_{t-1} in equilibrium.

The dynamic properties of this model depend on four parameters; a_{41}, g_1, g_2, and p. When $p = 0$, this model reduces to the initial model, (11.a) to (11.d), with consistency whenever $g_1 = g_2$ but stable only when $g_1g_2 < 1$. When p is greater than zero, the model is more complicated and mathematical solutions have not yet been obtained.

Numerical solutions have been generated which indicate that the model is stable when p is sufficiently large. These numerical results are summarized in Figure 4.13. In that figure, the hashed-in region in g_1g_2-p space is where explosive oscillations occur given a 10% rate of housing depreciation ($a_{41} = 0.90$). If the rate of depreciation is 2% ($a_{41} = 0.98$), the area of instability is extended to include as well the area between the two curves in Figure 4.13. We imply from these results that currently observable variations in the rate of housing stock depreciation have a very small effect on the stability of the model. Of much more importance is that job-seekers put enough weight (p sufficiently large) on housing considerations to overcome the oscillatory consequences of a concern for job forecasting. If, for example, 10% growth rates are

Figure 4.13. Stable Region for Model (13.a) to (13.f) given $a_{41} = 0.90$ and $a_{41} = 0.98$.



expected by both job-seekers and entrepreneurs ($g_1 g_2 = 1.21$), then a weight (p) of 0.175 or more on housing by job-seekers is all that is needed to ensure stability.

Although the model appears to be stable for reasonably broad ranges of parameter values, its consistency properties remain unknown. This is a problem to be the subject of a later paper.

4.8.6. Adaptive Job-Seekers and Urban Growth

Considering the final suggestion in section (d) above, we may redefine the model (11.a) through (11.d) as follows.

$$E_t = a_{10} + a_{11} g_1 P_{t-1} \qquad 1 > a_{11} > 0 \qquad g_1 > 0 \tag{14.a}$$

$$L_t = E_{t-1} + a_{11} p(P_{t-1} - P_{t-2}) \qquad p > 0 \tag{14.b}$$

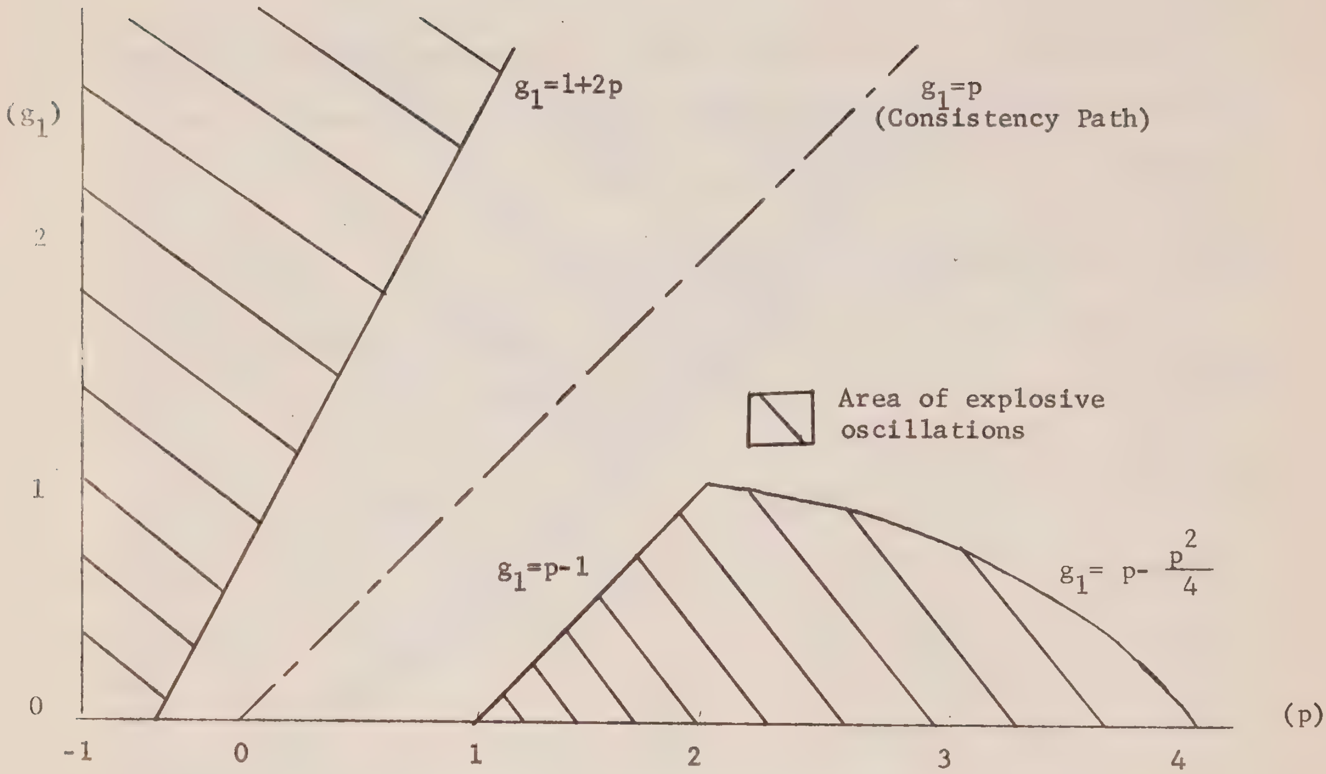
$$P_t = a_{20} + a_{21} L_t \tag{14.c}$$

$$a_{20} = - a_{10}/a_{11} \qquad a_{21} = 1/a_{11} \tag{14.d}$$

In this model, job-seekers consider the previous increase in population, $P_{t-1} - P_{t-2}$, make an expectation about the reproducibility of such a change, $p(P_{t-1} - P_{t-2})$, and estimate the new job offerings expected from that change, $a_{11} p(P_{t-1} - P_{t-2})$. By having job-seekers extrapolate from changes in population instead of absolute job offering levels as in (11.b), job-seekers have been made more adaptive to recent changes in growth.

The dynamic properties of this model depend on the parameters g_1 and p. Areas of the g_1 - p space which result in instability are displayed in Figure 4.14. Consistency can be shown to occur only where g_1 = p and it is noteworthy that consistency in this model always implies stability. The model is quite robust in that substantial deviations about the consistency line are required before stability is lost.

Figure 4.14. Consistency and stability in the Model (14.a) to (14.d).



Source: See text.

Consistency and stability are thus simultaneously possible within this model. Further, there are broad ranges of parameter values which result in stability in either a growth or decline process. This supports the argument that the present of job-seekers sensitive to recent population changes results in a stability and consistency not found in the initial model.

We have investigated three possible mechanisms to supplement the

original model, (11.a) to (11.d); variable participation rates, a housing sector, and adaptive job-seekers. The introduction of any of these mechanisms can be shown to introduce stability into growth processes for reasonably broad parameter regions. However, it has been suggested that the variable participation rate model may not be any more consistent than the original model. Since the consistency property of the housing model has yet to be established, it is concluded tentatively that the adaptive job-seeking model is the only one known to possess both stability and consistency.

4.8.7. Some Concluding Remarks

We began this section by asserting the hypothesis that the dynamics of urban economic growth are fundamentally dependent on the dynamics of the urban factor markets. A class of growth models, defined to be factor-passive, has been reviewed and the lack of any notion of factor markets has been noted. The concepts of a factor-recursive model and a semi factor-recursive model have been defined. Three examples of capital and labour recursive models have been presented. All three examples appear to suffer from a distinct lack of theoretical foundations.

In the remainder of the section on semi factor-recursive models, four new labour-recursive models are presented. In doing that, the object is twofold. First, the dynamic properties of some simple factor-recursive models can be studied. Secondly, the construction of such models enable us to examine more closely a plausible set of assumptions or theory which might underly them. Some small gains have been achieved on each count.

There is, however, a nagging aspect to the models constructed in this process and, by similarity, this also applies to Niedercorn and

Paelinck. Urban growth is usually thought of as a medium or long-run process. In our labour-recursive models, urban growth is determined by the short-run dynamics of the labour market. Is it reasonable to presume that the rate of economic growth over a long period is determined by a large number of seemingly unconnected and short-run labour market decisions? Perhaps it is. We might suspect however that there is also something else going on. We may point to the large number of empirical studies which have been notably unsuccessful in explaining migration solely in terms of short-run job-search and wage improvement. What else underlies labour migration, especially in the longer-run, is likely some variable which is totally unrelated to the short-run labour market. Models of long-run labour market behaviour must now be developed to complement the short-run models discussed in this paper. Then, urban growth can be related to both short and long-run behaviour and an appropriate emphasis derived for each.

The state of capital-recursive models is also unsatisfactory. We may distinguish here between fixed capital (buildings and equipment) and liquid capital (savings, liquidable assets). Fixed capital surpluses (deficits) may occur in an urban area if the area is declining (growing). In Bell's model, fixed capital surpluses (or deficits) have an effect on urban growth. In the short-run at least, fixed-capital recursive models have some theoretical justification. Conventional economic wisdom asserts however that liquid capital is perfectly mobile in a spatial sense. The usual theory does not allow for local investment to be tied to local capital sources given a well-developed liquid capital market. If liquid capital is converted to fixed capital at the place of best return, there can be no liquid capital shortage or surplus in the economic sense and no

factor-recursive mechanism. Ghali asserts that liquid capital is not mobile and that a capital-recursive model is appropriate. He does not however justify his assertion in any way.

We may pose the following questions about capital recursive models. Is a full (liquid and fixed) capital-recursive model possible? If so, what accounts for liquid capital immobility? Is this immobility related to city size or to other attributes of urban growth? How significant is capital immobility relative to labour immobility as an explanation of urban growth?

CHAPTER V

MARKOV DECISION AND CONTROL MODELS OF URBAN CHANGE

5.1 Introduction

To this point various descriptive models of urban dynamics have been presented. These models can of course have direct relevance to questions of policy as they are used for prediction and sensitivity analysis. We have seen that if the parameters of a Markov process can be made to be a function of policy variables, the model can be used to demonstrate the implications of alternative decisions. This section is concerned with carrying the analysis one step further, with the introduction of decision making process explicitly into the model. Thus, assuming purposive decision making, the problem becomes one of specifying a set of objectives to be attained or optimized and selecting the appropriate sequence of decisions (or decision rules) which optimize the attainment of these goals.

Typically in an urban content many people and agencies are continuously making decisions which affect the form and the functioning of an urban area and the interactions between and within such areas - decisions varying from the choices of a suitable place to shop by an individual to the decision to stimulate or discourage the growth of a major urban centre. The spectrum of decision making is indeed broad and varied. The decisions made at one level can condition the decisions that will be taken at other levels. In effect, certain decisions (e.g. assessment rates) set the parameters for the decision problem to be made by other actors in the urban system (e.g. choice of residence). Any attempt to capture in a single model the intricate

web of decision making processes in an urban area is certainly doomed to failure. Certain subsystems of decision making can be modelled, however.

This chapter attempts to investigate methods to model some of the dynamic aspects of decision making in an urban context. Extensive use is made of analogies from other areas of social science and operations research. Examples are presented which attempt to illustrate how such modelling could be undertaken -- many of these models as they stand are quite naive and certainly incompletely specified. They are presented to give a flavour of the type of modelling which must be undertaken to capture the dynamic nature of urban systems. Some of the modelling approaches, particularly optimal control theory, are really in their infancy as an operational models of systems as complex as cities. However, in certain areas of engineering and economics, control theory is an extremely vigorous field of research and in principle at least seems to come closer than any other approach as an effective way to investigate how dynamic feedback systems can be managed over time. It is increasingly important for those concerned with applications to monitor continually on-going scientific and technical research. With the acceleration of scientific activity, it is becoming increasingly possible for a widening of the gap between scientific developments and the means of solving real problems. If policy researchers are aware of new technical developments, it may be possible first to translate some real problems into an existing technical framework or secondly to influence the development of subsequent technical research so that it becomes more relevant to policy questions. In summary, then, optimal control theory and some of the other methods presented in this chapter may not be currently feasible ways of tackling urban problems in an operational way. They are presented here for two reasons: first, that they may soon, with some modifications, become feasible techniques and secondly, they may in their

current form give important qualitative insights into the dynamics of urban systems.

Markov decision models[1] are concerned with systems that can be described as a Markov process over which a decision maker can exert some control. In order to make it a decision problem, the concept of rewards or costs must be introduced. Associated either with the occupancy of a state (e.g. a land use type) or a change from one state to another (e.g. a shift of a job from one city to another) is a cost or reward. The goal of the decision maker is to minimize (maximize) the expected costs (rewards) of the system over some finite time horizon or if the process has no obvious terminal point the expected costs are averaged or discounted to the present using a suitable rate of interest.

The options open to the decision maker are generally in the form of some partial control he has over the rules of change, the transition probabilities, of the system. His problem then is to choose a way of selecting the probabilities so that he minimizes (maximizes) his expected costs (rewards).

Some simple problems of this type can be solved by enumerating all possible solutions; others require the use of linear, dynamic or other mathematical programming procedures. Still others necessitate the use of optimizing methods within a heuristic simulation framework where one is assured only of identifying approximately optimal solutions.

The remainder of this chapter considers the following topics. First, as an operational example of Markov decision processes, replacement models are considered, primarily focussing on the aging, maintenance, renewal and

---

[1]For a more formal introduction to Markov decision processes the reader is referred to Howard (1960, 1972), Derman (1972), or the relevant chapters in Hillier and Lieberman (1967) and Wagner (1969). The technical presentation here is incorporated into the substantive examples which follow.

replacement of urban housing. An intracity migration model is wedded to the housing model as an illustrative example of a larger scale model. The second topic concerns a re-interpretation of inventory models within an urban growth context. Next mathematical control theory is introduced in both its deterministic and stochastic forms. It is considered as a possible modelling framework to control certain aspects of an urban housing market, to ensure sustained urban growth, and to dampen fluctuations in a local economy. Finally, some notions of entropy maximization are introduced and applied to an urban growth problem.

In the presentation of all of these topics, the urban application are suggestive rather than definitive. All of the details, both of a conceptual and mathematical nature, have not been worked out. All of the approaches, however, appear to be promising enough to warrant additional research. Suggestions of such research are frequently made throughout the presentation.

5.2. Residential Change as a Markov Decision Process

One of the classic problems of operations research is the machine replacement problem. Very simply, a machine deteriorates physically with age and use; and with this deterioration, its efficiency decreases until it reaches a state where it is costly to operate or virtually useless and must be extensively overhauled or replaced by another unit. It is not unreasonable to assume that this aging process can be described by a Markov chain with the states of the system defined as the categories of machine condition. Given the nature of transition probabilities, which are ergodic under any plausible assumptions, the well known Markov chain statistics regarding state occupancy in intermediate stages of the process and in the steady state can be readily derived.

Associated with each transition or state occupancy is a cost or reward reflecting the machine's productivity as well as maintenance and replacement costs. Different policies will imply different probability and reward matrices. For example, a policy of preventive maintenance will increase the magnitude of the diagonal elements of the transition probability matrix and decrease the net payoffs in the reward matrices. Similarly a policy of early replacement will change the transition matrix

$$P_1 = \begin{bmatrix} .7 & .2 & .1 & 0 \\ 0 & .7 & .2 & .1 \\ 0 & 0 & .7 & .3 \\ 1.0 & 0 & 0 & 0 \end{bmatrix} \quad \text{to} \quad P_2 = \begin{bmatrix} .7 & .2 & .1 & 0 \\ 0 & .7 & .2 & .1 \\ 1.0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 \end{bmatrix}$$

and also the cost matrix.

One way to evaluate alternative policies is to weight the steady state probabilities by the costs of being in the corresponding state:

$$Z^K = \sum_{\ell=1}^{n} C_i^K \Pi_i^K .$$

This can be interpreted as the expected long run average reward associated with policy K. In many cases, of course, transient behaviour is of considerable interest and other performance criteria must be used.

A rather obvious analogy between the machine replacement problem and the housing renewal problem is apparent. Defining the states of the system as the condition of a housing unit, the aging and redevelopment process can be described as a Markov chain. A numerical example will perhaps be useful in fixing our ideas. Let P_1 and P_2 be the transition probabilities associated with two alternative policies.

$$P_1 = \begin{bmatrix} .7 & .3 & 0 & 0 \\ 0 & .7 & .3 & 0 \\ 0 & 0 & .7 & .3 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad P_2 = \begin{bmatrix} .8 & .2 & 0 & 0 \\ 0 & .8 & .2 & 0 \\ 0 & 0 & .8 & .2 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The second policy involves applying resources to the housing unit throughout the aging process in an attempt to retard deterioration. The reward matrices R_1 and R_2 reflect these differing costs and payoffs (rents).

$$R_1 = \begin{bmatrix} 4 & 3 & 2 & 1 \\ -12 & 3 & 2 & 1 \\ -16 & -12 & 2 & 1 \\ -20 & -16 & -12 & 1 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 3 & 2 & 1 & 0 \\ -12 & 2 & 1 & 0 \\ -16 & -12 & 1 & 0 \\ -20 & -16 & -12 & 0 \end{bmatrix}$$

(Note that since 9 of the 16 transitions are defined to be impossible, the corresponding elements in the reward matrices are not relevant but are included anyway for the sake of completeness).

The obvious question to the manager of a housing project is what policy or mix of policies should be adopted? One answer to this question can be found by formulating the problem as a dynamic programming problem. The famous "principle of optimality" relies heavily on the Markovian property of many dynamic systems. It simply states that an optimal policy for a N stage process must be such that the policy for the subsequent N-1 stage process resulting from the first decision is also optimal. Thus an N stage problem can be solved by solving N single stage problems stated in the form of recursive equations.

Returning to the example, we assume the terminal condition $V(o) = 0$. That is, if there are no stages remaining there is no payoff. (This assumption

of zero scrap value could be easily relaxed.) If there is one stage remaining we can evaluate the two alternative policies.

$$V^1(1) = \begin{bmatrix} 3.7 \\ 2.7 \\ 1.7 \\ -20 \end{bmatrix} \qquad V^2(1) = \begin{bmatrix} 2.8 \\ 1.8 \\ 0.8 \\ -20 \end{bmatrix}$$

If the housing unit is in first class condition (state 1), the expected reward under policy 1 (low maintenance) is

$$V_1^1\,(1) = \sum_{j=1}^{4} P_1{}^1{}_j\, r_1{}^1{}_j = .7(4) + .3(3) = 3.7$$

Similarly, the corresponding expected reward under policy 2 is

$$V_1^2\,(1) = \sum_{j=1}^{4} P_1{}^2{}_j\, r_1{}^2{}_j = .8(3) + .2(2) = 2.8$$

The other elements of the "value vectors" are calculated in exactly the same way. Thus it is clear that in the situation where only one state remains, policy 1 is better than policy 2 for all housing conditions except state 4 where the two policies yield the same result because they prescribe the same actions. Note that no matter how many stages remain these quantitites $V^1(1)$ and $V^2(1)$ represent the immediate expected reward associated with the two policies. Therefore to distinguish them from other expected value vectors we shall designate the immediate expected reward vectors associated with policies one and two as q^1 and q^2 respectively.

Consider now the process with two stages remaining. The expected reward $V_i^K(2)$ associated with policy k is composed of two components - q_i^k, the immediate one-stage expectation and the optimal expected reward for a situation with only one stage remaining. That is,

$$V_1^k(2) = q_1^k + \sum_{j=1}^{4} P_{1j}^k \; f_j(1)$$

$f_1(1) = 3.7$, $f_2(1) = 2.7$, $f_3(1) = 1.7$, $f_4(1) = -20.$

Thus,

$$V^1(2) = \begin{bmatrix} 7.1 \\ 5.1 \\ -3.1 \\ -16.3 \end{bmatrix} \quad \text{and} \quad V^2(2) = \begin{bmatrix} 6.3 \\ 4.3 \\ -1.8 \\ -16.3 \end{bmatrix}$$

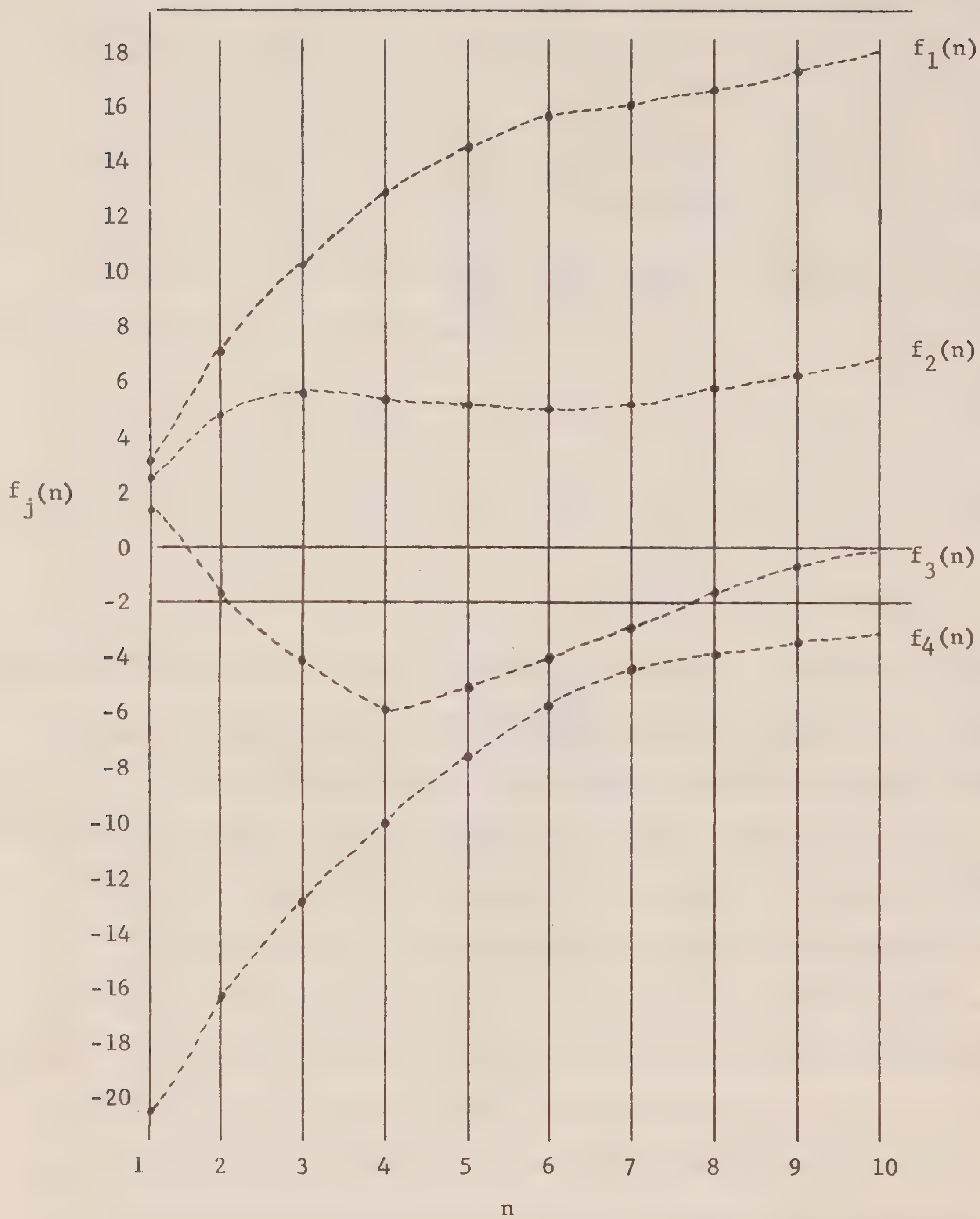
Therefore,

$$f(2) = \begin{bmatrix} 7.1 \\ 5.1 \\ -1.8 \\ -16.3 \end{bmatrix}$$

and the optimal policy is 1, 1, 2, 1 or 2 for states 1, 2, 3, 4, respectively.

Thus even with only two stages remaining, it is worth some maintenance to attempt to avoid entering the final state 4. Continuing in this way to solve for successively higher numbers of stages remaining, a graph of the results can be obtained (see Figure 5.1). Note how even with this simple example, the results are perhaps not trivially obvious. For example, the optimal value of the criterion functions are not monotonically increasing with time remaining for two of the initial states. The functions for a housing unit starting in either the second or third state behave in a more complex way than the other two. Consider the functions in turn. For state 4, an initial outlay of \$20 is necessary with certainty. Successive transitions have the effect of reducing this deficit as the new unit earns rent. Starting in state 3,

FIGURE 5.1


18
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
$f_j(n)$
$f_1(n)$
$f_2(n)$
$f_3(n)$
$f_4(n)$
1
2
3
4
5
6
7
8
9
10
n

the immediate positive expectation is quickly dampened by the need to renew (or replace) the unit quite soon. In state 2, the expectation is that revenue will accumulate for two or three periods, then decline as the facility is replaced and then rise as rents begin to accumulate again. Finally, a new housing unit increases monotonically, but at a decreasing rate. Although the range of expectations over different starting times narrows with increasing time remaining, the advantage of a good head start is of considerable importance even for a 10 period problem.

Concerning the <u>policy</u>, it is of interest to note that the decision rules are quite sensitive to the number of stages remaining. For some sequences, policy 1 dominates for some states and then policy 2 becomes dominant.

For ergodic Markov chains, a stable mix of optimal policies exists for infinite stage problem, i.e. a problem without a naturally definable end point or planning horizon. For the infinite stage version of this problem, the optimal decision vector is $D^* = (1, 2, 1, 1)$ or equivalently $(1, 2, 1, 2)$. Howard's (1960) policy improvement algorithm is used to derive this result. First, an arbitrary policy is specified and its value (average expected reward) calculated. Then it is determined whether this value is a maximum and if not how to improve upon it. This algorithm continues until the maximum value and optimum policy is determined. It should be noted that this procedure, as well as the previous value iteration one, is feasible for quite large problems. The number of states N is maybe a limiting factor in a few cases, but in general this is unlikely as we must only be able to solve systems of linear equations in N unknowns and for most cases data

availability will restrict us long before computational considerations[2]

Note a characteristic feature of both value iteration and policy improvement solutions. The sequence of decisions is not prescribed. Rather the rules for generating decisions as a function of the current state of the system are specified. This feedback type of solution is particularly suitable for managing systems over which the decision maker can exert only partial control. Future conditions are thus not known with certainty. Taking into account the probabilities of future states given the underlying stochastic mechanism and his limited ability to control the system, the decision maker implements a policy which optimizes his expectations, then observes the unfolding of the process, and adjusts his policy in light of the changing state of the system.

This problem as stated is of course quite artificial in that many rather obvious complications could be incorporated to make the problem more realistic. Two modifications are of particular relevance, however, not only to this housing problem but to virtually all dynamic urban control problems. We shall introduce these two extensions here with this admittedly rather simple residential renewal problems and return to both in other contexts. The first considers how to introduce horizontal structural interdependencies. The second is concerned with incorporating hierarchical controls on the system.

---

[2]The design of specific dynamic programming codes to solve particular problems can often be avoided by reformulating the problem in a linear programming format. This not only can reduce programming time (with an accompanying increase in computation time) but will also allow the modeller to have access to sensitivity analysis routines of linear programming (see Wagner, 1969 for more details).

Consider the situation where the rewards of one landlord are to some extent dependent on the actions of another. The dependencies may be either negative or positive, competitive or complementary. In the first instance, we have the situation where the landlords are competing for the same finite market, thus one entrepreneur gains at the expense of another. In the second case, the upgrading of one housing unit increases the general desirability of an area and thus increases profitability of improving other housing units. Both of these interdependencies manifest themselves spatially. Can such interdependencies be incorporated in the already developed dynamic programming framework? In order to facilitate the exposition let us make some simplifying assumptions. Let us assume that one of the housing units dominates the other in the sense that if both are in the "same" condition, unit 1 will earn a preferential rent. Decisions regarding the first unit can then be made independent of decisions regarding the second. In fact, we could assume that the preceding numerical solution pertains to the first decision maker's problem. Assume further that the second manager must choose between policies 1 and 2 or a mixture of the two. The difference in this case is that the reward matrices are functions of the changing condition of housing unit 1. For example:

$$R_1^{\,2} = \begin{bmatrix} 4 - \delta^1_{1i} \cdot a & 3 - \delta^1_{2i} \cdot a & * & * \\ * & 3 - \delta^1_{2i} \cdot a & 2 - \delta^1_{3i} \cdot a & * \\ * & * & 2 - \delta^1_{3i} \cdot a & 1 - \delta^1_{1i} \cdot a \\ -20 & * & * & * \end{bmatrix}$$

where $\delta^1_{ji} = 1$ if $j=i$, 0 otherwise

a = the amount of rent lost by unit 2 as a result of being usurped by unit 1.

* = indicates an irrelevant reward (i.e. transition defined to be impossible.).

A similar reward matrix could be defined for the second (preventive maintenance) policy.

Assuming that the second decision maker is aware of the policy and current state of the first unit, it is straightforward procedure to calculate the optimal policy of the second decision maker as a function of the number of stages remaining (in the finite horizon case), unit 1's current state, and unit 2's current state. Thus we can accommodate this simple form of interdependency at the cost of including an additional state variable. One should note that the number of calculations and the amount of storage necessary quadruples. Thus where before the optimal policy for the unit with n stages remaining was a 1×4 vector indicating what decision to make for each possible state, it is now a 4×4 matrix indicating the decision for each pair i, j of state i of housing unit 2 and state j of housing unit 1. Similarly, the information concerning expected rewards with n stages remaining is now a 4×4 matrix.

Without presenting the numerical results of such a problem, it is of some interest to note the sorts of behaviour such a model would be able to indicate. For example, it would certainly be to the advantage of the second entrepreneur to avoid as much as possible occupying the same state as the first since his returns are less under these circumstances. Thus, his policy will be to retard or accelerate the deterioration process so that his deterioration and renewal cycle is out-of-phase with the first. The case of multiple decision makers could also be considered using this same framework, but the number of state variables increase computational time and storage requirements exponentially. The case of mutual interdependence vastly

complicates the problem and resort would have to be made to game theoretic approaches, notoriously intractable for complex problems.

The second fundamental extension concerns the hierarchical control of systems. Many individuals and agencies are continually making decisions which influence the housing deterioration and renewal process outlined above. Often decisions are made at one level which essentially set the parameters for decisions which are made at another (lower) level. This is the case in the simple horizontal interdependence problem outlined above. The additional feature to be considered now is that the decision maker at one level is attempting to manipulate the parameters of the decisions process at the lower level in order to optimize the attainment of his goals. (In the previous case, the first decision maker only incidently affects the parameters of the second.) In the housing example, a public agency may be concerned with minimizing the expected average number of houses occupying the third and fourth condition categories. What policy (in the form of subsidies or penalties to the entrepreneurs) should the public agency adopt in order to influence the policies of the entrepreneurs so that expected occupancy of these states is minimized? For example, at what stage in the housing cycle should the government intervene to subsidize home maintenance and improvement? The public agency would undoubtedly have budgetary and perhaps other constraints on its actions.

This problem, although easily stated is not so readily solved. One approach would be to experiment with different reward matrices and by solving the dynamic programming problem for each set of matrices, attempt, by trial and error to move towards better and better solutions. A more systematic variant on this approach would be to identify an arbitrary subsidization - penalty policy; then solve the lower level problem for the associated reward

matrices; then optimally select government policy given this entrepreneurial decisi this cycle of solving sequentially for the two levels of decision making would continue until two successive iterations yielded no improvement. Whether or not the resulting government policy would be truly optimal depends in part on the specifics of the problem. More research needs to be undertaken on these convergence properties. If convergence to a global optimum is not assured, different initial policies should be used and the best of the local optima selected as an approximation to the global optimum.

This problem has only relatively recently been studied extensively. It is widely referred to as the decomposition problem (Dantzig, 1963 and Tintner and Sengupta, 1972). It relates to any situation where one set of optimizers is reacting to the rules of the game laid down by a decision maker at a higher level. This type of problem is particularly important in a partially controlled, decentralized economy. Dobell (1970) and Tintner and Sengupta (1972) emphasize this property in an economic development context, but it is equally important for urban systems. As Dobell (1970) states,

> . . . a decentralized system takes a great deal of control into its own hands and control therefore must be exercised relatively indirectly. . . . We really must specify our descriptive model with this individual optimizing behaviour . . . . built in as a part of the system the analyst has to take as given. Control is therefore implemented rather indirectly with possibly imperfect instruments. . . The formulation may entail study of much higher order systems, with control exercised through influence on intermediate variables of little interest in themselves.

This multilevel decision making approach is a far truer representation of most real urban policy situations than ones which assume either a monolithic central planning agency making all the decisions at one extreme or the perfectly competitive model at the other extreme which assumes many independent, unregulated small decision makers. Unfortunately, as is often

the case with increased realism, increased computational difficulties are encountered. The problem of decomposition and decentralization decision making is certainly a vigorous area of research in applied mathematics and promises to be of great relevance to many hierarchical urban decision problems. We shall return to this problem in the discussion of control theory below.

The number of basic maintenance-replacement alternatives has been limited to two in the example. One could and certainly should expand this list to a larger number to reflect quite subtle changes in policy. It is extremely important not to preclude certain policies because of a simple oversight. For example, replacing or renewing the unit before it reaches the fourth state should perhaps be considered. Also, state 4 could be declared absorbing in which case the unit would be removed from the housing stock, presumably by being sold. Also both the maintenance costs and off-diagonal transition probabilities for a "renewed" house may be more than that for a new house. Thus a comparison of the options of renewal vs. new construction could be made in terms of their respective long run expected costs.

Such additions to the options to be considered would necessarily increase the number of calculations, but linearly rather than exponentially. (One approach would be to solve the problem initially considering a small number of quite different policies, and then solve again by considering policies only slightly different from the optimal policies of the first solution.)

Inclusion of other policy options would be particularly important in the example of two interdependent decision makers. An initial policy encouraging rapid or very slow deterioration would enable the second entrepreneur to reach a steady state condition out-of-phase with the first housing manager very quickly. Instead losses would be made up in subsequent stages as each

unit would in most cases earn its maximal possible rent.

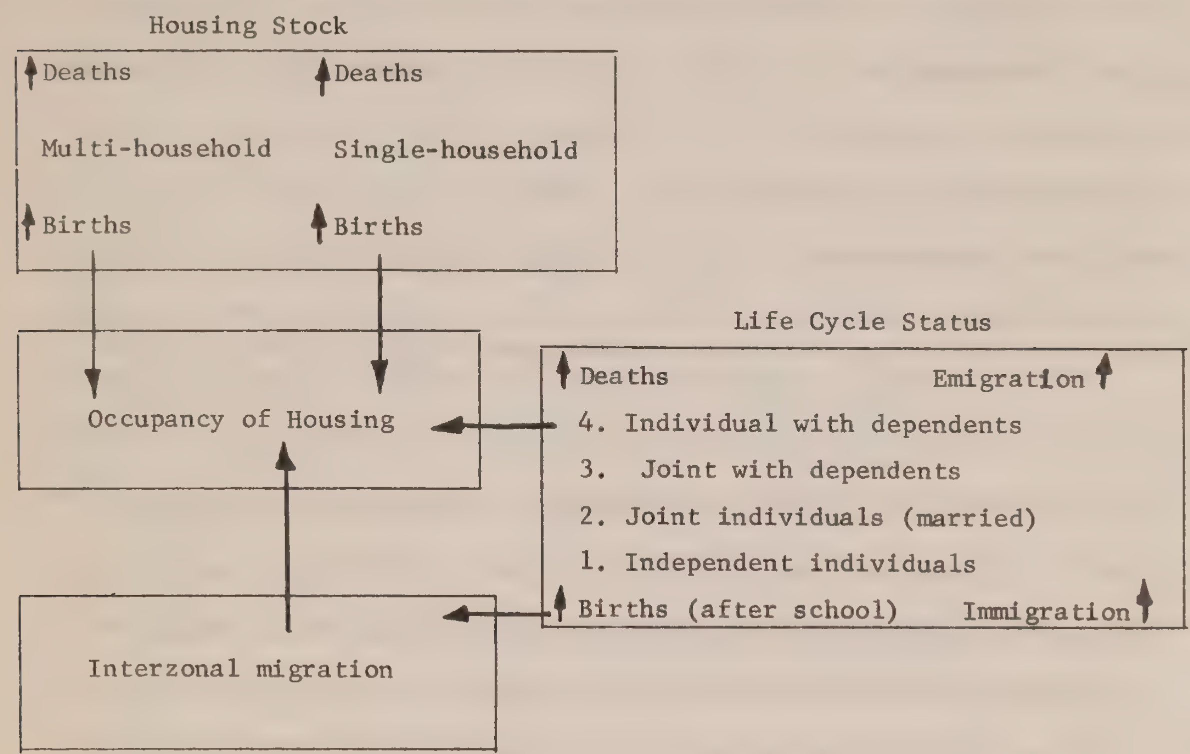
An empirical study has already been described in an earlier section which addresses itself to a housing renewal problem using Markovian analysis (Wolfe, 1967). Although similar, one should note that because it does not use a dynamic programming framework, Wolfe's approach is rather a myopic solution - solutions which are lucrative in the short run are not necessarily optimal over a longer period. Dynamic programming has the characteristic that it does attempt to anticipate the long run implications of current actions. It would be interesting to know whether or not Wolfe's type of solution would yield a better description of developer-landlord behaviour. It certainly could not yield a more profitable solution.

In addition to the cycle of investment and deterioration many other dynamic forces affect the urban housing market not the least of which are migration of households within, from and to urban areas and the social-demographic process of the life cycle, Migration processes are of course to a large degree influenced by the life cycle. Let us now consider a more elaborate housing model which attempts to incorporate all of these dynamic features of an urban system.

Consider Figure 5.2. which is part of a possible larger set of stocks and flows comprising the interdependencies of a city. We are attempting to combine a housing occupancy model based on the personal life cycle, a housing stock model and an interzonal migration model. Types of housing are to be provided in the different zones of a city which will maximize some objective function in the long run, such as maximum profit from real estate development. As well as being interesting in itw own right, this particular example is designed to show the capabilities (and limitations)

of Howard's method of choosing optimum paths in a Markov process.

FIGURE 5.2.



We may commence with the life cycle status of the inhabitants in a zone. We can expect this to be relatively constant throughout a city although such an assumption is not absolutely necessary. Life cycle states are defined so that a matrix of transition probabilities may be constructed. Since the states themselves define the probabilities of requiring a house or an apartment, we can obtain the probabilities of changing types of housing (Table 5.1).

TABLE 5.1. LIFE CYCLE STATUS IN A REGION

	Probability of Requiring	
	House	Apartment
Independent individuals	.20	.80
Joint individuals (married)	.35	.65
Joint individuals with dependents	.80	.20
Individuals with dependents	.65	.35

Life cycle status is built from demographic tables and is arranged to be Markovian ergodic and thus can be run to the steady state. Run in unit time periods, it may be used to change housing change probabilities.

Consider migration probabilities from region i to region j. It would almost certainly be best to make these a function of other attributes of the system, including life cycle status and particularly place of employment and any change therein, but we shall take them as given. In behaviour terms it is sensible to describe the probability of an individual arrival in a region coming from a particular background. On the other hand, it is not very sensible behaviourly to describe the probability of an individual departure from a region to a particular region. The departing individual might be indifferent between many regions and does not assign probabilities to each.

We require migration probabilities as dependent variables, i.e. the probability of an individual arrival in a region originating from a

particular background. This will involve a description of regional characteristics in terms of an individual's status. A region's ability to attract migrants from different regions and social status can be regarded as demand for that region's housing. Demand in Region A will be a function of the life cycle status in Region B plus the probability of moving from B to A (relative to moving from A, C, D, etc.).

In order to manage the housing stock we conceive a rather abstract operator representing a collective entrepreneur seeking his maximum economic advantage. However he has very long vision and seeks the long-term gain. While the optimum he obtains is his maximum profit, it is also a social optimum if the structure of housing prices be accepted as mirroring real social forces. Since the long term profit is maximal, the demand over a long term is being met where it is greatest. If a planning agency has to accept the same assumptions as the entrepreneur, then using dynamic programming in such a model would guard against its seeking immediate rewards. There is no reason why prices should not be composed to provide a purely social welfare ranking, i.e., allow greater profit for a low income house or apartment than a high income house or apartment.

The operator sells to a household from Region J = 1, 2, 3 which previously had a certain type of housing and he acquires this unit in exchange plus a cash settlement. The new occupant is trading his existing residence with the supplier of the new unit, paying or being paid a certain sum to balance. Assume he can borrow from bank as required so that he is thus always within the system. In providing a vacant residential unit in Region I we require the probability that the new occupant will come from Region J and so vacate a unit there.

New entries into the system do provide problems. An immigrant may be

allocated between regions on the basis of previous regional choices and between housing types on the basis of income and life cycle status. The probabilistic allocation of a particular unit depends on the propensity of households to migrate to that region for that type of housing. In the case of a sale to immigrants, new housing units are added so that increases in population moved cause increases in housing stock, either (a) intensifying occupancy of land or (b) extending lands occupied. The former implies an increased desire for apartment dwelling and thus a greater proportion of apartments. It would only result from a differing status of people, e.g. younger without dependents or having foreign tastes. The latter probably but not necessarily means an increasing proportion of single family housing. In order to pin things down, consider Table 5.2.

TABLE 5.2.

States		
Region 1	Policy	
Low income house: (H.L.)	(e)	retain as is
	(f)	demolish and build new apartment unit-low income
	(g)	demolish and build new apartment unit-high income
	(h)	upvalue to high income house
High income house (H.H.)	(e)	retain as is
	(f)	demolish and build new apartment unit-low income
	(g)	demolish and build new apartment unit-high income
Apartment unit	(e)	retain as is
Low income (A.L.)	(g)	upvalue to high income apartment unit
Apartment unit High income (A.H.)	(e)	retain as is
Additional units for immigrants to system (I)	(i)	build new low income house) only for regions
	(j)	build new high income house) with vacant land.
	(k)	build new apartment unit-low income
	(l)	build new apartment unit-high income

The states of the system are four types of housing unit in N regions, the demand for which depends on the existing interregional pattern of migration, the life cycle status and the economic level. For each type of housing there are various alternative policies available to the real estate operator. Note however that if only the "retain as is" policy is used it would be possible to find the demographic steady state solution corresponding to the existing housing stock.

Table 5.3 sets out the headings of the full matrix.

TABLE 5.3.

Region 1, 2, 3, . . .

do

State Region J	Policy	Region 1				Region 2				Region 3				Immigrants inc. rest of Canada
		HL	HH	AL	AH	HL	HH	AL	AH	HL	HH	AL	AH	
1. HL	e													
	f													
	g													
	h													
1. HH	e													
	f													
	g													
1. AL	e													
	g													
1. AH	e													
1. I	i													
	j													
	k													
	l													
2. HL	e													
	f													
	g													
	h													
2. HH	e													
	f													
	g													

The means for generalizing this model are fairly obvious. It would be fairly simple to include a model of employment change by individual manufacturing and commercial plants which would handle structurally forced job changes and another of 'normal' non growth or decline, labour turnover both of which models would feed into the housing occupancy sector. At this point we could almost specify a commuting model linked to a land-use model.

It can be seen that there is one enormous advantage in settling for a Markovian dynamics. Once this simple form is accepted, one can concentrate on trying to see how the different parts of the urban system fit together and affect each other. In other words, one can give one's whole attention to describing the system. The question of linking processes which are individually Markovian into n dimensional matrices is difficult. Thus, for example, in an additive process, such as a reservoir, if inputs (and outputs) are independently random, the level of the water can be modelled as a Markov chain. However we could not then use this feature as input to a further chain because the levels are autocorrelated. It is possible however to use a double categorization of states, i.e. for a reservoir with n units and previous input of m units in time t, the transition probability to n + u units of storage and u units input is such and such. It is possible to obtain the overall probability of storage levels, but this would clearly be doubly Markov for use as an input to a further chain. However, it is possible to link chains by increasing the categorization of states, provided that we adjust all the phenomena to the same time scale. This also suggests that the fact that transition probabilities which, when estimated in real life often appear to change from period to period, could in fact be the result of the imposition of a

simple process on a more complicated one.

However one must ask what does the steady state mean in terms of residential urban land use? If the economy is static, presumably land use is quite static and transition probabilities would then have no meaning. Presumably then the transition probabilities refer to periods of expansion. It is possible that the degree of expansion may not matter much since, when the rate is low, probabilities for change will be scaled down and those for the status quo increased. Consequently results would be the same with a variable time scale. However, it is unlikely that the transitions would be unchanged in periods of contraction.

One of the drawbacks of the Howard approach is that costs and returns are fixed independently of the system. It is thus possible to find an optimal strategy which allows the decision maker to do better than others in the system. Thus in Howard's taxicab situation, a particular policy in waiting for a customer in each town allows the dynamic programmer to achieve maximum profits. But if many taxicab owners followed the same policy, presumably the matrix he chooses in each town would have a different return matrix associated with it. Waiting time at the cab stand would go up so that the desirability of such a policy will go down. Because he can make more than average profits, presumably the system is not designed efficiently so that his choice of policy presumably produces an increase in social welfare. However, as more taxi cabs follow this policy, then the profitability of this policy will go down and that of the others go up. Presumably maximum social welfare occurs when taxicabs are indifferent between policies. Nor can we expect that the behaviour of consumers would be indifferent to the policies adopted: the need to travel between towns will not change but as fewer cabs cruise and more in cab stands, it seems likely that more users will walk to the

nearest stand so that the alternative policy transition matrices will change. Nor need balanced equilibrium in behaviour occur: the history of travel via private automobile vis a vis public transport in North America often shows an equilibrium without public transport existing. Equally the cruising cab could disappear.

Obviously, we cannot criticize the decision process literature for not handling all of these problems. It is being developed to show how particular policies should be chosen by the individual in particular cases - that these policies should also be socially optimal is not really relevant. However, if these approaches are to be used by governments catering for all the actors in the system then it seems clear that the returns matrix will need to be more complex. In the taxi cab case, for example, it would be necessary to be more explicit about the returns matrix which would cause drivers to change strategies sufficiently so as to be indifferent between policies. In a sense, dynamic programming of Markovian trajectories reveals the extent of the disequilibrium occurring in terms of price theory.

We have already shown that on a computational level, there is a relationship between dynamic programming and linear programming. The latter is often a more convenient method for solving problems originally formulated using a dynamic programming format.

A related problem has been developed by Eastman and Kortenek (1970). This model considers the complex interdependencies between two sets of decisions - the provision of housing and schools of certain types and qualities. Certain types of housing are more likely to attract one type of household (life cycle status) than other types. These households, again with certain probabilities change their life cycle status, and

therefore their schooling needs over time. How can the housing developer and the school board phase their decisions of housing and school mixes so that neither schools nor houses of any type are grossly under utilized? The authors formulate a rather ingenious and intricate solution to this problem combining Markov chain and linear programming models.

Finally, a class of models very closely related to these approaches is called recursive programming. Generally couched within a linear programming framework, these models simply use parts of the solutions generated in time t as parameters in the subsequent time period. Thus, in the several agricultural examples to which these models have been applied (see Day, 1963 and McDonald, 1972, for example) the pattern of land use in time t imposes a constraint arising from questions of inertia on the pattern of land use in the following time period. Schlager (1965a,b) in an urban development application constrains land development by partially endogenous predictions of demands for land - predictions based on past system behaviour. This rather simple synthesis of an optimizing model with autoregressive approach makes this quite an attractive modelling framework.

5.3. Inventory Models and Urban Change

This study has made extensive use of models which were originally designed to solve quite different problems. Using analogies can be both a powerful means of devising ways to solve or gain insights into problems and also extremely dangerous in that misleading and incorrect conclusions can be drawn because of the lack of perfect isomorphism between the two problems. This section represents an unabashed exercise in analogous modelling. We look at some models originally designed to optimally manage inventory and ordering (or production) levels over time and attempt to translate them, sometimes with minor changes, often with none, into models which are relevant to some urban issue. Sometimes the analogies are somewhat strained and artificial, but they represent a first attempt to unite a powerful set of techniques with some very pressing urban problems.

Consider an urban area about which the government has some concern. It wishes to prevent this city from losing population. The desirable population level is P. Assume a constant job multiplier of α (representing the families of workers and the purely population - serving workers and their families). Thus the desirable number of jobs is $S = \dfrac{P}{1 + \alpha}$. Although P is the desirable population level, the government is willing to tolerate lower levels p. Associate with p is the basic job level $s = \dfrac{p}{1 + \alpha}$. Thus when the number of jobs falls to or below s, the government is committed to inducing or creating jobs so that the number of basic jobs is increased to S and population to P. (Of course this instantaneous multiplier is a simplifying assumption which could easily be relaxed). The size of the difference S-s may reflect the government's indifference between population levels over that range and also the

lumpiness of the costs associated with inducing or creating new jobs.

Assume further that jobs are leaving according to an independently random, time invariant process, e.g., Poisson. Superimposing the deterministic government policy on the random process of job departures yields a simple Markov chain model where the states of the process are the number of basic jobs (or population levels).

For example, if departures are Poisson with $\lambda = 2$ and the policy is $(s,S) = (95, 100)$, then the following transition probability matrix would describe the job (and population) fluctuations of the city:

	95 or less	96	97	98	99	100
95 or less	0	0	0	0	0	1.0
96	.865	.135	0	0	0	0
97	.594	.271	.135	0	0	0
98	.323	.271	.271	.135	0	0
99	.143	.180	.271	.271	.135	0
100	.053	.090	.180	.271	.271	.135

Note that N^2 (= 36 parameters in this case) are readily derived from three, one of which λ is empirical while the others s and S are prescribed by the policy maker.

Mean first passage times, limiting state probabilities and other Markov statistics may be calculated from the transition probability matrix to give a complete statistical picture of the expected job and population fluctuations together with variances.

The limiting state probabilities could be particularly useful in the evaluation of government policy. Suppose there is a penalty attached to

being in each state. Occupying state 1 ($X(t) = 95$ or less) is particularly costly since it implies government expenditures for job creation. Other states 2, 3, 4, and 5 may also have costs in that the city has something less than desired population. Assuming such cost estimates are available, the long run average cost per unit time (that is, after the system is in equilibrium) is

$$\sum_{j=1}^{6} C_j \, \Pi_j \qquad \text{where presumably } C_6 = 0 \quad \text{and where}$$

$\Pi = (\Pi_1, \Pi_2, \ldots \Pi_6)$ is the limiting state probability vector.

Concerning cost component C_1, departures in any one period may be such that the number of jobs dips to levels considerably below 95. The subsequent costs to bring them up to 100 will vary with the size of this discrepancy. C_1 must be calculated by weighting the different types of cost (creating 5, 6, 7, 8, . . . jobs) by the probabilities of the occurance of these discrepencies, thus obtaining the expected cost associated with occupying state 1.

Of course, by varying the assumptions (the values of λ, s, or S) different stochastic processes could be generated. Assuming S and λ fixed, s could be varied. Associated with each s_k would be a different transition probability matrix the statistics for which could be generated along with long run expected average costs.

Given the data on the costs and the dynamics one might ask the question what is the optimal value of s_k - i.e. the job level below which the government will not tolerate. This question may be answered using methods of complete enumeration, linear programming or dynamic programming.

Little need be said regarding complete enumeration. We simply experiment with every feasible value of s_k and choose the one for which long run expected costs are minimized.

For large problems this method is not feasible. Manne (1960) and Derman (1962) have developed linear programming methods to solve such problems. In the above example, if the system is in some non-optimal state j, the planner can do two things: create or induce jobs S-j or allow the system to proceed and take action at some later period. Associated with each policy is an expected cost C_{jk}. For purely technical reasons we assume that when in state j, the planner will enact policies according to a to-be-determined probability distribution. (As it happens, the probability distribution is degenerate so that in state j, the same decision is always taken.) The linear programming formulation is then:

Minimize $$\sum_{j=1}^{n} \sum_{k=1}^{k} X_{jk} C_{jk}$$

Subject to $$\sum_{k=1}^{k} X_{jk} - \sum_{\ell=1}^{n} \sum_{k=1}^{k} X_{ik} P_{ij}^{k} = 0 \qquad j = 1, 2, \ldots, n$$

$$\sum_{j=1}^{n} \sum_{k=1}^{k} X_{jk} = 1$$

$$X_{jk} \geq 0 \qquad j = 1, 2, \ldots, n \qquad k = 1, 2, \ldots, k$$

Where X_{ik} (the solution variable) is the long run probability of occupying state k and taking decision k.

(In the present example we have assumed only two decisions are possible - do nothing or create S-j jobs. By introducing additional solution variables X_{i3}, X_{i4}, etc., other possible policies could be included in the optimization

process - creating somewhat fewer than S-j jobs, in an attempt to determine the optimal value of S as well as s.)

The actual optimal decisions can be obtained from the equation:

$$D_{jk} = \frac{x_{jk}}{\sum_{k=1}^{k} x_{jk}}$$

where $D_{j1} = 0$ and $D_{j2} = 1$ or $D_{j1} = 1$ and $D_{j2} = 0$ for all j in our example.

Under plausible assumptions regarding the cost components C_{jk}, the value of s would be the highest value of j such that $D_{j2} = 1$ where k = 1 implies that S-j new jobs are created.

Long run solutions may not be entirely relevant for two reasons. The planning horizon may be relatively short because the government needs to demonstrate its effectiveness over the next three to five years, not the next twenty or thirty. Secondly, long before the process reaches an equilibrium, the parameters governing the system may change. Thus, transient behaviour of the system over the next few years may be more important than long run steady state behaviour. If this is the case, the value iteration method of dynamic programming (Howard, 1960) is more relevant. The optimal policy is then determined as a function of the current state and number of years remaining in the planning period.

Howard's policy improvement method of dynamic programming may be relevant for long run decisions where the initial state is known. In this formulation both transient and equilibrium behaviour is optimized.

Of course, the above model could be readily adapted to the case of a rapidly growing city where S represented an <u>upper</u> limit of city size. Any growth beyond S would be stopped by encouraging (subsidizing) or forcing firms to relocate to other cities at a certain cost (direct as well as

indirect). s would then be regarded as an ideal size of the town. Assuming that jobs were arriving according to a Poisson or other known random process, the resulting increases and decreases of town size could be described according to a simple finite Markov chain.

These two models could be generalized to account for the case where both arrivals and departures occurred according to time invariant stochastic processes. For example, suppose jobs depart according to a normal distribution with mean 2 and variance equal to 4. This would be similar to the Poisson ($\lambda = 2$) assumption of the previous numerical example. In this case, even though the dominant trend would be of decline, negative departures (i.e. arrivals of jobs) would occur with a certain positive probabilities. With a specified (s,S) policy a finite Markov chain model could easily be derived. Further, given some additional information on costs, the optimal value of s could be derived using linear or dynamic programming methods.

A further modification of the model could involve the relaxation of the assumptions regarding constant s, and S. Suppose, for example, the target population S changes through time. The format described above, with minor modifications, could be used to determine the optimal sequence of decision rules which would "phase down" (or alternatively "gear up") economy of an urban area over time. This would yield a non-stationary Markov chain. Thus limiting behaviour would be neither possible to study nor of interest. Transient, finite horizon methods would be used.

This modelling framework, like the inventory models upon which it is based, is quite passive in nature in that it accepts as given the parameters of the departure (and arrival) processes. In the case of the "naturally" declining urban area, the government is essentially pouring water into a

deep and leaky well. The policy consists of temporary, remedial measures, nothing is being done to stop the leaks. Similarly, in the case of a "naturally" increasing centre, the mean ratio of arrivals is given, periodically to be halted, but then to be allowed to increase to its natural level.

One could modify the model still further by postulating that one or more of the parameters of the process is a function of a government policy variable. For example, $\lambda(t) = \delta - bG(t)$, that is, the mean number of departures is a negative function of government subsidies or expenditures on public services. Rather than responding to the outcomes of the "natural" process, this strategy would have government policy incorporated into the process itself. While from the expenditures point-of-view, this may seem quite similar to the former approach, it reflects a quite different philosophy of control.

We note in passing that within this same framework we could readily specify s and S as functions of time - non-stationary tolerable ranges $(s(t), S(t))$.

Another extension would imbed this basic inventory model into a larger context which includes intercity relationships. Thus far, it has been assumed that jobs arrive and depart according to some probabilistic process, but we have not taken into consideration of the origins or destinations of these arrivals and departures. Obviously the departure of a job from a city is often associated with the arrival of that same job in another city (in addition to the disappearance of the job altogether); similarly the arrival in one city is also a departure from another (again in addition to the generation of completely new jobs). How can we accommodate such multicity interdependencies and what are their possible policy implications?

Collins (1970) has modelled the intercity movement of jobs as a simple Markov chain. Assume that the intercity movement data have been organized so that we have two intercity transition probability matrices - (1) a forward matrix P composed of elements P_{ij} , which are the probabilities of a job entering city j given it is going to relocate from city i; (2) a backward or reverse matrix $\hat{P}$ made up of elements $\hat{P}_{ij}$ which are the probabilities of a job coming from city i given it has just arrived in city j. Note that in both instances only intercity movements are considered. Thus in the first case, for example, the probability is conditional in two ways - the current location is known and it is also known that an intercity relocation is about to be made.

Of course in addition to these two matrices we have a matrix for each city indicating the probabilities of increasing or decreasing size. These probabilities are determined as in the simple city case by inherent "death" and "birth" processes in addition to a specified (s,S) or (S,s) policy.

Analytical solutions could be calculated for small systems but for larger ones simulation methods would have to be used. Considering the problem in a simulation format is perhaps helpful to gain an appreciation for the complex systems interactions which would have to be followed, even with this quite simple formulation. In a three city system in which two towns (1 and 2) are declining and one town (3) is growing, the probability that the third city will experience a unit increase in number jobs is a function of (1) the aggregate growth rate of the system and how that growth is distributed among the cities (2) the probability that town one will lose a job multiplied by the transition probability that that job will relocate to town (3) (P_{13}) plus the product of the same two probabilities for town 2

minus the joint product of these four probabilities. If the beginning of the second stage number of jobs in town 3 is greater than or equal to the specified upper limit S_3, the job level will be reduced to s_3. These jobs will be distributed to towns 1 and 2, according to probabilities P_{31} and P_{32}. The system is allowed to continue, the planner making adjustments whenever job levels fall to or below s_1 and s_2, on to or above S_3. To consider the other case of jobs falling below s_1, new jobs are induced to town 1 so that the new job level is S_1. These $S_1 - s_1$ or more jobs come to city either by the creation of entirely new jobs or by inducement of jobs from cities 2 and 3. In the job migration case, it is clear that if P_{31} is large, the system adjustments complement one another - job levels in city 1 are maintained by dampening the increases for the growing centre. If on the other hand P_{21} is large, the system adjustments exacerbate the problem. City 1's job level is maintained by relaxing the job levels in city 2. This increases the probability that city 2 will need remedial action. If P_{12} is also large, the adjustment process compounds the problem still further. It is quite likely that these probabilities are high in actual situations. Both cities are declining and therefore are likely to be characterized by similar economic structures. In addition, there are probably marginal firms in these towns which are likely to be responsive to government inducements - far more likely than those in the prosperous growing centre. For both these reasons, we would expect the probabilities P_{12} and P_{21} to be high relative to probabilities P_{32} and P_{31}. Government subsidization would have the effect of increasing the velocity of movement between the declining centres, incurring significant costs with no comparable benefits. Similar results would hold in a larger urban system with two or more growth centres. Inforced departures could

have the effect of increasing the number of movements between growth centres rather than to the declining centres.

Of course government policy is likely to be more sophisticated, giving subsidies only to desirable moves (from growing to declining centres). This would necessitate the estimation of two sets of intercity transition probabilities depending on whether the move was to be "natural" or induced. The estimation procedure in the induced case could be rather simple setting probabilities of undesirable moves to zero and increasing the other non-zero probabilities proportionately so that row (or column) sums were equal to unity.

This proliferation of transition probabilities occurs even in the simplest multicity case. Intercity probabilities are not constant even though they are mover probabilities. The intercity probabilities must be adjusted when one or more cities is at its critical upper or lower state. Under the assumptions of the model, no out migration would be possible during this "regeneration" stage of a declining centre's development. Note that the adjustments to the probabilities require no additional data collection, simply additional calculations. These adjustments would make a significant difference only when the city in question represented a large component of intercity migration or when several cities happened to reach their critical job levels simultaneously, an unlikely event. Alternatively we could assume out migration possible with the same probability, but that new job level was a random variable with an expected value of S rather than known with certainty.

This deterministic feature is perhaps the most questionable aspect of the approach at both single city and multicity levels as it has been presented to this point. It is not realistic to view government policy in encouraging

and discouraging the growth of certain centres as equivalent to placing an order for a certain quantity of goods. Even in a centrally planned economy, this representation is not adequate. As it happens, it is the easiest assumption to drop. The government may be able to prescribe approximately the job level it desires and sets up a policy which has this level as an expectation with a known probability distribution. Thus the three state transition matrix for a declining centre might be

$$\begin{bmatrix} .1 & .2 & .7 \\ .3 & .7 & 0 \\ 0 & .3 & .7 \end{bmatrix}.$$

Subsequent analysis of this and other similar transition matrices could proceed in much the same way as the deterministic policy case.

Let us look at the inventory modelling framework as a way to analyze numerous urban capital projects such as public housing, office buildings, and transportation. Such projects tend to be "lumpy" or discontinuous because of the economies of scale arising from larger projects. Unlike inventory models, however, these systems are typically non-stationary with supply and demand tending to increase over time. Such processes are easily, in principle at least, made stationary by redefining the state of the system to be "excess capacity." Figure 5.2 could thus represent a typical pattern of construction and depletion of excess capacity through time. Such patterns are typical of inventory model solutions. Periodic shortage in supply are accommodated in this formulation; associated with these shortages are costs (loss of efficiency and or consumer (or voter) displeasure).

Qt = amount of new construction in time period t.

St = shortage (unsatisfied demand) in period t.

Mt = excess supply over demand at start of period t.

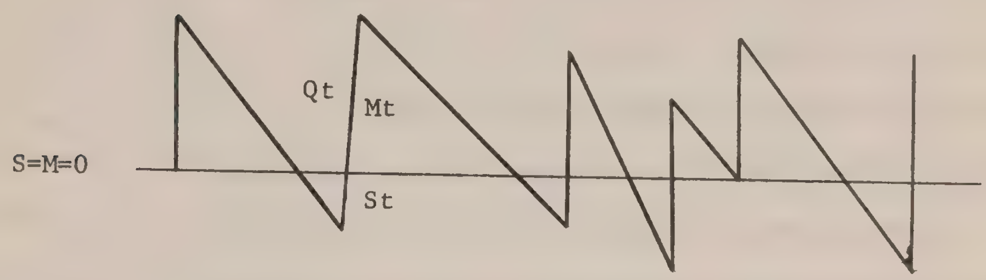


Figure 5.2.

There are thus three types of costs involved - production costs, holding costs, and costs of shortages. The simplest assumption regarding production costs is that they are linearly related with size of project with each project encurring a fixed "set up" cost K no matter how large. Alternatively convex or concave production cost functions could be assumed. "Holding" costs refer to those costs associated with maintaining one unit of excess capacity per unit time. This is the cost of maintaining and operating offices, buildings, housing projects, buses and trains which are being used at less than their full capacity, or more generally and realistically perhaps, houses and other facilities earning less than maximal rent. As units of this excess capacity are eliminated, operating costs are at least covered by revenue. Any excess of revenue over costs could of course be included as a negative cost. For transportation and other public provided goods, revenue and cost calculations are not undertaken in this way, but the

principle remains the same.[3] The cost functions associated with "holding" and "shortage" need only be monotonically increasing.

Such a re-interpretation of a simple inventory model could be used to analyze and, in a restricted sense, optimize a sequence of construction projects. Because of the lumpy nature of supply and the continuous increases in demand, the system will quite naturally pass through a sequence of cycles of excess supply and demand.

In the simplest formulation $r(t)$, the requirements in the t^{th} period are known deterministrically. These requirements are a function of increasing demand over time and the depletion of existing stocks of capital as they depreciate over time.

Well known algorithmic procedures are available to minimize the sum of production, holding and shortage costs over finite or infinite time horizons. (See Veinott, 1966, for a review.) Stochastic versions of this same basic model could be presented. Also, lead times, very important in capital projects, could be incorporated without excessively complicating the model. Objective functions which include the cost of large changes in production from one time period to the next (production smoothing) could be considered. But instead, let us briefly turn to the problem of how the model could be modified to deal with spatial interdependencies within urban areas.

Veinott (1966) reviews some attempts to model multi-product and multi-locational inventory systems. Many products are close substitutes

---

[3] For example, the users of an expressway do not pay directly for the privilege of using them; thus revenues from that project do not exist or at least are difficult to identify. This complicates the inventory model application, but does not invalidate it.

for each other and shortages in one may be made up by supplies of another. Similarly, the unavailability of a product at one warehouse may be compensated by the availability of the same product at a more distant distribution point. In the urban case, office space in one part of the city is at least a partial substitute for similar space in other areas of the city. Deficiencies in one area can therefore be made up by excess supplies in other areas. Transportation facilities, either different routes or different modes, can be thought of in this same way. Thus the optimal regional production schedules and consequent cycles of excess supply and deficient demand would be to some degree out-of-phase with each other through time similar to the pattern illustrated in Figure 5.3.

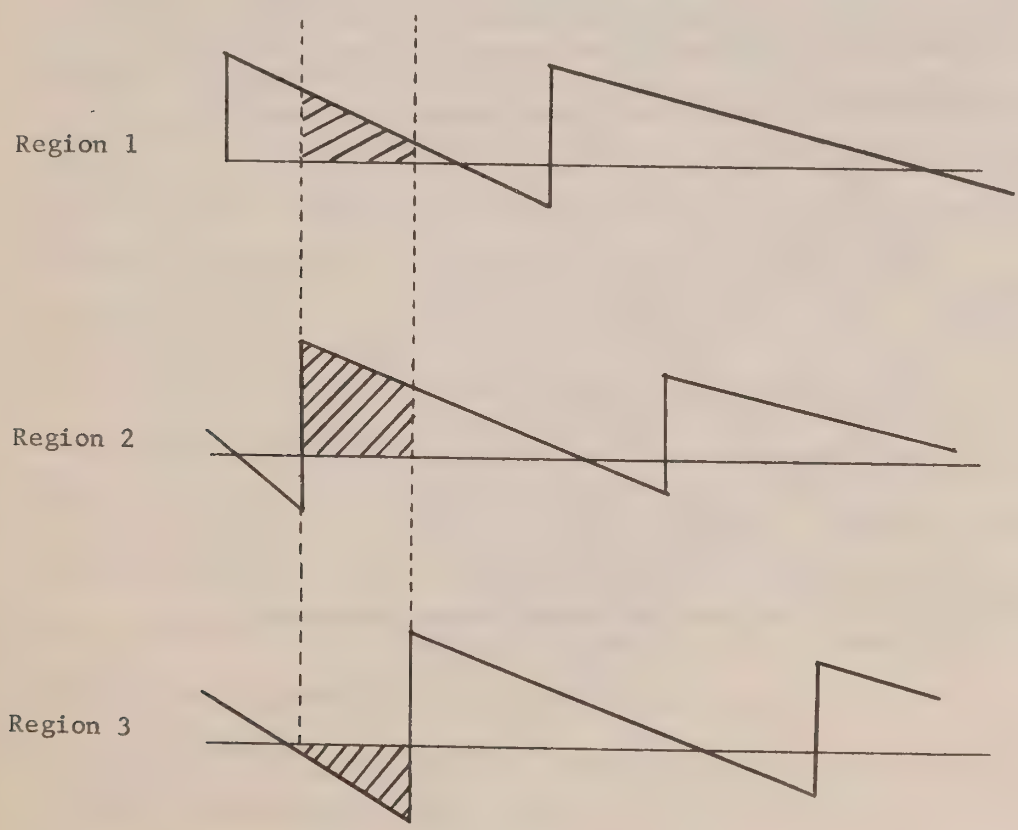


Figure 5.3.

The system taken as a whole may no longer have supply deficiencies, and again taken as a whole may reduce the total production necessary to satisfy demand.

Erlenkotter (1967, 1972) places such a problem within a dynamic programming framework. For a two facility problem, the state space is described by a vector (Y_1, Y_2) indicating the surplus capacities at the two locations. Assuming a constant increase in demand over time and no aggregate backlogging of demand, surplus capacity depletion continues until no surplus capacity exists. This depletion is characterized by a movement from initial state along a 45° line in Figure 5.4 until the line representing no surplus capacity is reached. Here investment is made in one of the facilities (a vertical or horizontal shift in Figure 5.4. In this formulation the linear programming transportation problem minimizes distribution costs at each point in time given capacities, demands and

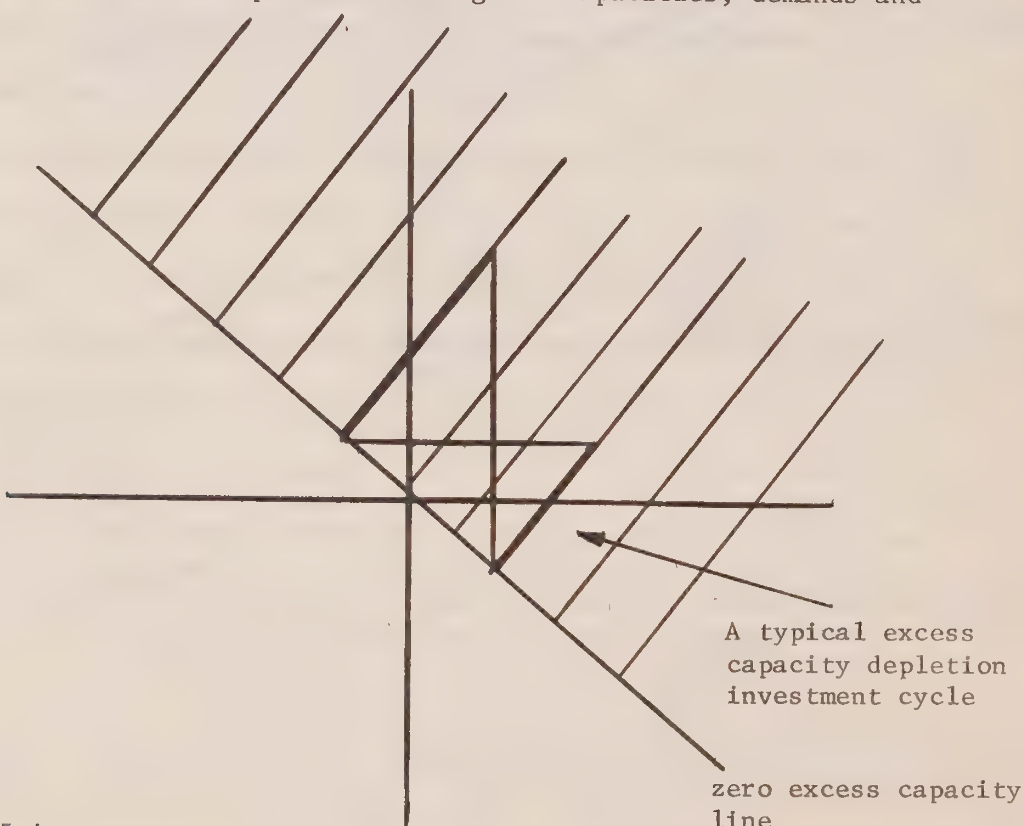


Figure 5.4

per unit transportation costs. This is assumed to be given as $P(Y,t)$. More formally, the dynamic programming functional equation is:

$$C(Y^h) = \underset{x_i \, \epsilon X(Y^h)}{\text{Min}} \{ \int_o^{\tau\, hk} P(Y^h,t) \; e^{-rt} dt + e^{-r\tau} \; f(x_i \; D) + \int_{\tau \; hk}^{\Delta \; hk} P(Y^{ht},t) \; e^{-rt} \; dt + e^{-r\Delta hk}(C(Y^h)) \}$$

$$\text{for all } Y^h$$

Thus the program minimizes (1) the integral of discounted transportation costs up to the time of expansion τ_{hk} plus (2) the discounted expansion costs plus (3) the discounted expansion costs from the time of expansion up to the end of time interval $_{hk}$ plus (4) the future costs beyond Δ_{hk} (Erlenkotter, 1972). These and similar models are summarized in Sheppard (1973).

All of these models have stochastic as well as deterministic versions. Generally the stochastic versions investigate the process where demand is independently random with known parameters. A relevant variant of this deals with an inventory system in which supply as well as demand is stochastic. The problem in this case is to exercise whatever controls are available by in essence selecting from alternative probability distributions to minimize the expected value of some performance criterion.

The literature in inventory theory is exceedingly rich - one of the first class of management problems in which mathematics was successfully and elegantly applied. (See for example Arrow, Karlin, and Scarf, 1958) It has been successful in terms of the quantitative results and qualitative insights it has given to some management problems. It is not unlikely that such results and insights may be forthcoming in urban contexts.

The following section is a generalization, formalization and extension of some of the topics just considered.

5.4. Generalized Markovian Decision Processes

We have already discussed the difficulty of reconciling the usual form of growth mechanism with the planning of radical changes in the urban structure. Almost any form of positivistic approach is likely to sanction the existing and condemn what is not. Since we almost surely require some contagious mechanism so that a large centre means a viable centre, this line of reasoning holds out little hope for ameliorating the lot of small towns and poor regions. However, we were at pains to point out that, unlike biological species, locational decisions are not made on the basis of existing conditions but rather on a forecast of what future conditions will be. Normally, the forecast is dependent on what the present is like but this need not necessarily be the case. Thus although the usual style of time series modelling would not encourage the possibility of deliberately "growing" a city, it should be possible to operate directly on expectations.

The forecast can be a guarantee from a government that in twenty years, say, there will be a large diversified city in a place where presently one does not exist. For many firms this diversification may be too far off in time and too vague in specification. However there must be many firms which are modestly footloose and which would settle for a guarantee of diversification in the future, especially if this were linked with a pleasant site for living and perhaps some temporary financial assistance.

The main drawback to this approach is that the entrepreneur must exchange a future environment which is highly likely on the basis of the present for one which is highly likely only on the basis of a governmental promise. This means, of course, that the policy must either be made

enforceable on governments or that it be acceptable to all parties. There must be a determination to meet the promise by all possible means including carrot, stick and the relocation of existing or new government facilities. If obligations are fully accepted and entrepreneurs convinced that plans will be realized then in fact there may not need to be much governmental action at all. Clearly the policy would be applied not only to manufacturing industry but to all sectors: a fairly large city must be in operation within a period of ten years.

Obviously such projects are major undertakings and could not be scattered around the landscape. If the government actually bought the land, and in secret so that it was cheap, a large part of the cost of the project could be realized from its sale or rent. Equally, such a city should be located to take over the role of a high order central place for servicing the surrounding area.

However, we still need some form of model to guide policy implementation and which incorporates a 'normal' mechanism which can be influenced by guaranteeing an end result. The main factor reconciling the normal and the planning mechanisms is that growth is based on expectations in both cases. In addition, the stochastic nature of the process is quite similar to the general epidemic model described earlier. We seek to explore the manner in which a government could develop a large city by 'guaranteeing' that it would be in existence by a certain date. The model for planned development is certainly inadequate but the method of attack is interesting.

The present section is an outline of de Leve's (1970) work on generalized Markovian decision processes, in itself an extension of Howard's (1960) and Jewell's(19 studies and uses Weeda's (1970) practical example. For each initial state, the

evolution of the system is described by a homogeneous strong Markov process and if a decision maker does not intervene it is called a natural process. If he does, there is an instantaneous transition in the state of the system defined by the probability distribution of the state into which the system is transferred. In order to have a decision occur at each point in time, the null-decision is introduced which simply allows the system to follow a natural process between intervention points with its initial state being that existing after an intervention.

For each state x within the state space X of the system there is a set D(x) of feasible decisions, d, based only on the present state of the system. A strategy z assigns a decision d to a state x. The result of the natural process and a strategy is called a decision process: this is a homogeneous strong Markov process.

With only costs involved (not discounted) the criterion for an optimal strategy is the expected average cost per unit of time.

We assume that the government has two weapons. The first is outlay on promotion, advertising, etc. and the second is the 'drafting' of tame enterprises - presumably governmental units. The latter would only be used in emergencies when the minimum planned development was not being met. Let us assume that firms are attracted in such a manner that the departure from planned employment is a Poisson process. Depending on where employment is relative to the target figure, promotion levels can be altered to change the flow of firms. If there is a drop below some minimum target level, government drafts occur to keep it at that level. We may also conceive a maximum gain above target which is useful in 'causing' a change in promotion levels but may not be necessary.

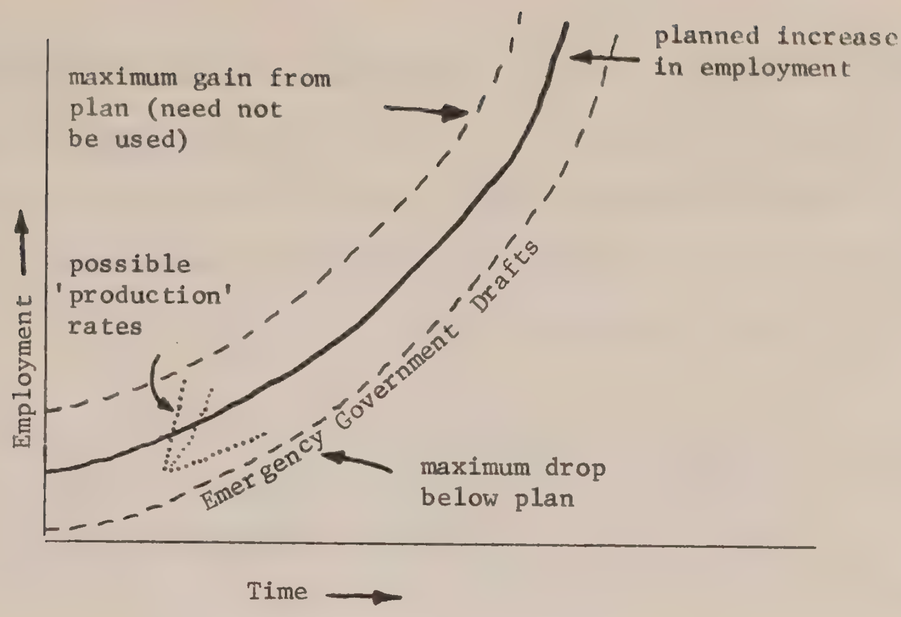


FIGURE 5.5.

The following costs are involved:

(1) promotion costs $c_p(i)$ per month for level 1_i, $i = 1, 2, . . ., N$ with $c_p(0) = 0$.

(2) costs $c_q(i,j)$ of switching over from level i to level j.

(3) costs c_r per employee for government drafts.

(4) target attainment cost c_s per employee per month. These will be negative (i.e. they are benefits) and are introduced to prevent the optimum path being close to the 'maximum drop below plan' line.

Continuous promotion levels are finite, 1_i, $i = 0, 1, . . ., N$ with $1_o = 0$. The maximum gain above target is M. The state of the system is specified by two state variables, the promotion level, i, and the attainment level, s. The state space

$$X = \{x = (i,s) \mid 0 \leq s \leq M; \quad i = 0, 1, . . . , N \}$$

The natural process with initial state (0,s) assumes a zero promotion level and involves the generation of employment relative to the plan and the first

emergency draft when $s = 0$ after which the state (0,0) is permanent. With initial state (i,s), $i > 0$, there is a fixed promotion level i. If s reaches 0, an emergency draft occurs to keep it at 0. A decision d in the state (i, s) leads to a transformation to the state (j,s') so that a set of feasible decisions in (i,s),

$$D(i,s) = \{ (j,s') \mid j = 0, 1, \ldots, N; \; s' = S \}$$

Each strategy requires an intervention in state (0, 0) and in states (i, M), $i = 1, \ldots, N$. At (i, M), the transformation is to (0, M). Hence the set

$$A_o = \{(0,0)\} \cup \{ (i, M) \mid i = 1, \ldots, N\}$$

/does not intersect

Generally let A_o be the set of states say A_{o1} and A_{o2} (referring to costs and time) reached from an initial set via a natural process. It may be reached from state x via a natural process using a random walk $\underline{w}_{o1}$ ($\underline{w}_{o2}$) incurring expected costs $k_o(x)$ and having an expected duration $t_o(x)$. Alternatively it may be reached from x following an intervention at x; the walk is now $\underline{w}_{d,1}$ ($\underline{w}_{d,2}$) with expected costs $k_1(x;d)$ and duration $t_1(x;d)$.

Let $\quad k(x;d) = k_1(x;d) - k_o(x)$

and $\quad t(x;d) = t_1(x;d) - t_o(x)$

The walks are identical if d is a null-decision. Hence in this case,

$$k(x;d) = t(x;d) = 0$$

The functions do not depend on any particular strategy and need be determined only once.

In our problem, we let

$$A_{o,1} = A_{o,2} = A_o$$

so that

$$\underline{w}_{o,1} = \underline{w}_{o,2} = \underline{w}_o$$

$$\underline{w}_{d,1} = \underline{w}_{d,2} = \underline{w}_d$$

the duration of a $\underline{w}_o$ walk having (i,s) as initial state is a random variable $\underline{t}_i(s)$ with expectation $t_i(s)$; the associated costs are $\underline{k}_i(s)$ and $k_i(s)$. From (0,s) its end state is $(0,0) \in A_o$; from (i,s), $i > 0$, its end state is (i,M).

Generally, for strategy z and initial state x, let $\{ \underline{I}_n = \underline{I}_n(z;x),\ n \geq 1 \}$ be the sequence of <u>future</u> intervention states. It can be shown that this sequence constitutes a discrete homogeneous Markov process in A_z. Let $p^{(k)}(A;z;x)$ be the probability that $\underline{I}_k$ belongs to A. There is a $q(A;z;x)$ for which

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p^{(k)}(A;z;x) = q(A;z;x)$$

For each $x \in X$, there is a stationary probability distribution $q(A;z;x)$ which satisfies

$$q(A;z;x) = \int p^{(1)}(A;z;y)\, q(dy;z;x)$$

Furthermore, if x_1 and x_2 belong to the same ergodic set $q(A;z;x_1) = q(A;z;x_2)$

$\lim_{T \to \infty} \underline{k}_T(z;x)/T$ represents the average costs per unit time for each initial state $x \in X$.

They are equal to

$$g(z;x) = \int_{A_z} k(I;z(I)\, q(dI;z;x) / \int_{A_z} t(I;z(I))\, q(dI;z;x)$$

for each x belonging to an ergodic set, and

$$g(z;x) = \int_{A_z} g(x;y)\, q(dy;z;x)$$

for each x from a transient set. In this latter case, the average cost per unit time is a random variable with expectation g(z;x) instead of being constant.

The functional equations for $t_i(s)$ and $k_i(s)$ are now derived for the case that $i = 0$. Let $\underline{\tau}$ be the time interval between the start of the walk and the arrival of the first deficit. Let $\underline{y}$ denote its size. If $\underline{y}$ exceeds the allowable deficit $s > 0$ an emergency draft is made and the walk thus ends. If $\underline{y}$ does not exceed the allowed tolerance then the walk continues from state $(0;s-\underline{y})$ and from then on the duration will be $\underline{t}_o(s-\underline{y})$. It follows that

$$\underline{t}_o(s) = \underline{\tau} + \begin{cases} \underline{t}_o(s-\underline{y}) & ; \quad \underline{y} < s \\ 0 & ; \quad \underline{y} \geq s. \end{cases}$$

Since $E\,\underline{\tau} = 1/\lambda$, the expected duration

$$t_o(s) = \frac{1}{\lambda} + \int_o^s t_o(s-y)\, F(dy); \qquad 0 < s \leq M$$

During time interval $\underline{\tau}$, stockholding costs amount to $c_s s\, \underline{\tau}$. If $\underline{y} \geq s$, then the costs of the necessary emergency purchase, $c_r(s-y)$ have to be added. If $\underline{y} < s$, then costs $\underline{k}_o(s-\underline{y})$ are incurred.

$$\underline{k}_o(s) = c_s s\, \underline{\tau} + \begin{cases} k_o(s-\underline{y}) & ; \quad \underline{y} < s \\ c_r(\underline{y}-s) & ; \quad \underline{y} \geq s \end{cases}$$

The expected costs

$$k_o(s) = \frac{c_s s}{\lambda} + c_r \int_s^\infty (y-s)\, F(dy) + \int_o^s k_o(s-y)\, F(dy); \quad 0 < s \leq M$$

If the initial state is $(i, M) \; \varepsilon A_o$ with the system producing at a level $i > 0$ so that stocks should rise but cannot, then the walk ends. Then,

$$t_i(M) = 0 \qquad ; \quad i = 1, \ldots, N.$$

$$k_i(M) = 0 \qquad ; \quad i = 1, \ldots, N.$$

The duration

$$\underline{t}_i(s) = \begin{cases} \dfrac{M-s}{\ell_i} & ; \quad \underline{\tau} > \dfrac{M-s}{\ell_i} \\[2ex] \underline{\tau} + \underline{t}_i(0) & ; \quad \underline{\tau} \leq \dfrac{M-s}{\ell_i} \; ; \; \underline{y} > s + \ell_i \underline{\tau} \\[2ex] \underline{\tau} + \underline{t}_i(s+\ell_i \underline{\tau} - \underline{y}) & ; \quad \underline{\tau} \leq \dfrac{M-s}{\ell_i} \; ; \; \underline{y} \leq s + \ell_i \underline{\tau} \end{cases}$$

The expected duration in functional form,

$$t_i(s) = \frac{M-s}{\ell_i} \int_{(M-s)/\ell_i}^{\infty} \lambda e^{-\lambda\tau} \, d\tau + \int_o^{(M-s)/\ell_i} \tau\lambda e^{-\lambda\tau} \, d\tau$$

$$+ \, t_i(0) \int_o^{(M-s)/\ell_i} \lambda e^{-\lambda\tau} \, d\tau \int_{s+\ell_i\tau}^{\infty} F(dy)$$

$$+ \int_o^{(M-s)/\ell_i} \lambda e^{-\lambda\tau} \, d\tau \int_o^{s+\ell_i\tau} t_i(s+\ell_i\tau - y) \, F(dy)$$

$$0 \leq s < M$$

$$1 \leq i < N$$

Differentiating with respect to s,

$$\frac{dt_i(s)}{ds} = \frac{\lambda}{\ell_i} t_i(s) - \frac{1}{\ell_i} - \frac{\lambda}{\ell_i} t_i(0) \, (1-F(s))$$

$$- \frac{\lambda}{\ell_i} \int_o^s t_i(s-y) \, F(dy) \qquad ; \quad 0 \leq s < M$$

$$1 \leq i < N$$

In a similar way, we obtain

$$\frac{d\,k_i(s)}{ds} = \frac{\lambda}{\ell_i} k_i(s) - \frac{c_p(i)}{\ell_i} - \frac{c_s S}{\ell_i} - \frac{\lambda}{\ell_i} k_i(0)(1-F(s))$$

$$- \frac{\lambda}{\ell_i} \int_o^s k_i(s-y)\, F(dy)$$

$$- \frac{\lambda c_r}{\ell_i} \int_s^\infty (y-s)\, F(dy) \qquad ; \quad 0 \leq s < M$$

$$1 \leq i < N$$

The expected duration and expected cost associated with the $\underline{w}_o$ walk from initial state $(i,s)\ \epsilon X$ are obtained from the previous equations.

The criterion function $g(z;x)$ is determined using the functional equations in $r(z;x)$ and $c(z;x)$.

$$r(z;x) = Er(z;\underline{I}_1) \qquad x\ \epsilon\ X \tag{13}$$

and

$$c(z;x) = k(x;z(x)) - r(z;x)\, t(x;z(x)) + Ec(z;\underline{I}_1) \qquad x\ \epsilon\ X \tag{14}$$

where $\underline{I}_1$ is the first future intervention state when x is the initial state. Hence,

$$Er(z;\underline{I}_1) = \int_{\Lambda_z} r(z;I)\, p^1\, (dI;z;x)$$

It can be shown that (13) and (14) have a solution and for each solution

$$(r(z;x),\ c(z;x)) \quad \text{we have} \quad r(z;x) = g(z;x) \qquad x\ \epsilon\ X$$

By adding to (13) and (14) the condition

$$c(z;e_j) = 0 \qquad j = 1, \ldots, m$$

The resulting system of equations has a unique solution. By means of $r(z;x)$ and $c(z;x)$ an iteration procedure can be given which converges on the minimum expected average costs per unit time from the initial state, x.

5.5. Mathematical Control Theory: An Introduction

Control theory is concerned with the optimal management, regulation, or control of dynamic systems. Thus, in a very real sense the Markovian decision models already considered are an integral part of control theory. For historical reasons more than anything, control theory has tended to be associated with a particular type of dynamic optimization problem and solution procedures.

Earlier we have considered systems of linear difference equations such as $x(t) = Ax(t-1)$ where x is an n-dimension state vector and A and n x n time invariant matrix of coefficients which relate the current state of the system to immediately preceding states. This is the assumption of first-order Markovian dynamics. (Higher order dynamics can of course be introduced at the costs of increasing the number of state variables.) Such a model can be used to describe and analyze the behaviour of urban systems often yielding quite complex and surprising results. The solutions to systems of difference equations are relevant to policy in that they predict behaviour over time and the sensitivity of behaviour to variations in initial conditions and parameters may be determined. If these variations can be related to alternative policies, these models can be very useful in identifying appropriate decisions.

Control theory carries the relevance to policy one step further by incorporating decision making into the model itself. In general then the linear first order dynamics are

$$x(t) = Ax(t-1) + Bu(t-1)$$

where x and A are as before and u is a vector of controls variables. For example, u_i could indicate the level of government expenditures in the i^{th}

city. B then is a time invariant matrix which relates this state of the system x to the control variables. For example, the diagonal elements b_{ii} could indicate the sensitivity of the variable value in the i^{th} city to expenditures in that city whereas b_{ij} ($i \neq j$) would indicate the spill-over effects (positive, negative, or zero) of expenditures in one city on other cities in the system.

The control problem is to determine the sequence of controls or decisions $u(o), u(i), \ldots, u(T)$ such that a criterion function is optimized and the dynamics of the problem are satisfied.[4] The criterion function is some specified function of the states of the system $x(1), x(2), \ldots, x(T)$ which is to be maximized or minimized. It may be a function of the terminal state of the process $x(T)$, e.g. we wish to maximize the number of houses of a given type available for occupancy ten years from now. Alternatively, the criterion function may be related to transient behaviour of the system e.g. the number of houses of a given type which are available over the next ten years. The criterion function is often in the form of a quadratic cost function.

In addition to the dynamics and criterion functions, there may be constraints placed either on system behaviour or the control variables. These constraints may be physical in nature (e.g. there cannot be a negative number of houses) economic (expenditures cannot exceed some specified budget β), or relate to objectives which are not included in the objective function (the number of houses in each year must be greater than some specified number).

---

[4] This is perhaps the simplest form of the control problem. In addition for many social systems, the control interface B and process dynamics A can be controlled to effect changes in system behaviour (Bailey and Hott, 1971).

Finally the dynamics may be augmented to include variables which are not controllable either directly such as u or indirectly such as x. The dynamics of the linear first order problem are then

$$x(t) \;=\; Ax(t-1) + Bu(t-1) + CZ(t-1)$$

where Z is the vector of exogenous variables. In the housing example, Z could include anything not controllable by the decision maker from the cost of lumber to the size of the population. In general, the inclusion of a vector Z with a known probability distribution is the way in which the control problem is made stochastic.

Solution procedures depend upon the specific nature of the problem, in particular the mathematical form of the dynamic, the criterion function and the dynamics. If all of these equations (and inequalities) are linear, then the well known methods of linear programming may be used. More generally, however, non-linear programming, dynamic programming, the calculus of variations and Pontryagin's mazimization principle are used.

One of the few applications of control theory to urban problems concerns the housing market (Weathersby, 1970a, 1970b). Concerning the decision to invest in and provide rent for public housing facilities, Weathersby introduces the following state variables:

$DU_i(t)$ = the number of low income dwelling units in region i in period t.

$P_i(t)$ = the low income population in region i in period t.

$LNC_i(t)$ = the number of new low income dwelling units which were begun in region i last year (this lagged control variable is included to accommodate the assumed two year construction lag in a lag-one model).

The control variables are:

$NC_i(t)$ = the number of low income dwelling units in region i whose construction was initiated in period t.

$RS_i(t)$ = the rent subsidy in dollars granted in region i in period t.

The exogenous variable is:

$R_i(t)$ = the average market rent for low income housing

Using these variables and some quite plausible assumptions, the dynamics are written for the two region case as follows:

$$\begin{pmatrix} DU_1(t) \\ DU_2(t) \\ P_1(t) \\ P_2(t) \\ LNC_1(t) \\ LNC_2(t) \end{pmatrix} = A \cdot \begin{pmatrix} DU_1(t-1) \\ DU_2(t-1) \\ P_1(t-1) \\ P_2(t-1) \\ LNC_1(t-1) \\ LNC_2(t-1) \end{pmatrix} + B \cdot \begin{pmatrix} NC_1(t-1) \\ NC_2(t-1) \\ RS_1(t-1) \\ RS_2(t-1) \end{pmatrix} + C \cdot \begin{pmatrix} R_1 \\ R_2 \\ 1 \end{pmatrix}$$

Some of the elements A, B, and C are estimated statistically from past system behaviour, others are based on common sense and subjective judgment.

A non-linear utility function is postulated which attempts to weight the relative importance to the decision maker of new dwelling units and population levels in each region. The problem is then solved subject to certain budgetary and political constraints using Pontryagin's maximum principle. Time paths for each of the state variables are derived under different budgetary assumptions and these are compared to the solutions with no controls. Similar to the dual variables of linear programming, this solution procedure yield co-state variables or multipliers which can be interpreted as the imputed marginal utilities of each of the state variables at each stage of the process. The marginal utility of a state variable in time period (t, T) is equal to its current marginal utility plus the marginal utility resulting from its current value in all future periods t + 1, t + 2, . . ., T.

Similarly, as a by-product of this solution procedure marginal utilities of control variables and constraints. Weathersby argues as others have before him that these imputed values in effect simulate a market situation in a context where none really exists. He suggest that in principle at least a central planner's objectives as incorporated in his criterion function could be implemented on a decentralized basis by means of these artificial internal pricing mechanism. The ability to decentralize planning decisions in an effective manner should be recognized as crucial in urban planning situations.

In another paper also concerned with the provision of housing, this time by the private sector, Weathersby (1970b) presents various formulations of the dynamic process of filtering, the way in which people improve their housing by moving up to more costly housing thus leaving behind a vacancy will allow yet another family to improve its housing. Assuming this step by step improvement occurs over a time interval rather than instantaneously Weathersby models the developer decisions regarding what price levels, at what quantitites and when to construct housing where here to there is a time lag between the decision to build and the completion of the housing unit. The criterion for these decisions is profit maximization. The developer is assumed to have no social objectives whatsoever. The solution to this multistage decision problem would certainly be of interest to a developer. It is also of interest to a public agency in that it allows it to "look over the developer's shoulder" and thus anticipate developers decisions. It can also be used to determine the sensitivity of residential development to alternative policies. By subsidizing development (or changing the tax and interest rates) the private sector will respond in certain ways and the population will move to housing of better quality according to its

availability and household preferences. Considerable insight may be gained by experimenting with alternative policies and observing the developer's reactions as they attempt to maximize their profits. Even more interesting and ambitious is to look at the two-level optimization problem of setting the parameters (per unit profit levels) which will maximize the attainment of some social goal (e.g. minimize the number of households in the lowest price level, presumably substandard housing). That is, what are the values of the parameters which will force the private developers to behave in a socially desirable fashion, as they attempt to maximize their profits? This is resolved by applying decomposition principles of mathematical programming. In this way optimal plans can be synthesized rather than an arbitrary set evaluated. Of course, the data and technique are currently insufficient to solve all but quite simple problems, but it seems clear that such techniques will be increasingly adopted and revised to clarify and in some instances to resolve actual urban problems.

Bailey and Holt (1971) discuss the decentralization and coordination of urban decisions within the context of Mesarovic's multi-level, multi-goal systems. In an admittedly simple but revealing example they postulate two agencies - one concerned with housing construction the other with factory construction. The housing sector constructs Y_1 million square feet per year according to the equation

$$Y_1 = m_1 - u_1$$

where m_1 as the quantity of resource inputs

u_1 represents disturbances due to competition in factor markets from the industrial sector

$$u_1 = Y_2 - \tfrac{1}{2} m_2$$

The industrial sector produces Y_2 million square feet of factory capacity per year according to the equation

$$Y_2 = m_2 - 2u_2$$

where, as before m_2 is the quantity of resource inputs used by this sector

and $u_2 = Y_1 - \frac{1}{2}m_1$ represents disturbances due to competition from the other sectors.

Assuming each sector has a goal of constructing one million square feet per year at minimum cost, each might have a criterion function

$$\text{MIN } G_i = m_1^2 + (Y_i - 1)^2$$

whereas the overall system goal would be

$$\text{MIN } G = G_1 + G_2$$

If each agency acts independently, ignoring the interdependencies U_i, then they select that level of inputs to $m = (\frac{1}{2}, \frac{1}{2})$.

But this is certainly suboptimal in terms of G. To improve performance, a coordinator C_o is introduced which modifies the goals of the agencies (via taxes, subsidies or simply changing the rules of the game):

$$G_1^1 = G_1 + B_1u_1 + \frac{1}{2}(B_2 - 2B_1)\, m_1$$

$$G_2^1 = G_2 + B_2u_2 + \frac{1}{2}(B_1 - B_2)\, m_2$$

where

$$B = (B_1, B_2) \text{ is the coordinator parameter.}$$

By means of an iterative procedure of message exchange the coordinators interact. C_i sends its desired $m_i\, u_i$ to C_o which then sends out a new vector B. If each C_i uses rules such that at step S in the iteration $[\, m_i\ u_i\,]_s$ is such that it minimizes G_i with respect to the previous

coordinator parameter and the coordinator C_o uses the rule:

$$B_s = B_{s-1} + [u(B_{s-1}) - Km(B_{s-1})]$$

where $u = [u_1 \; u_2]$

$m = [m_1 \; m_2]$

and K is a simple functional relation which shows the actual u obtained for each given m.

If this procedure is followed, then it converges to a set of messages

$$B = [-8/5, \; -2]$$

and $m = [\frac{1}{2}, \; 2/5]$.

Trinkl (1973) provides another example of hierarchical resource allocation decisions solving the problem by making the central decision maker allocate a specific level of departmental resources and a set of revised values of specific objectives to account for the interdependencies between agencies. Sage and Smith (1973) analyze the decomposition problem within the context of an urban dynamics model, but entirely for purposes of increasing computational feasibility. Developing procedures whereby public agencies or public and private systems can act with a large degree of independence but still proceed along paths which are desirable from a broader social point-of-view is a problem on which a very high order research priority must be placed (Mesarovic, 1970). It is of particular relevance within the context of an urban system which is embedded within a multi-level governmental structure and an economic system in private sector is an important component.

Another set of issues arising from control theory relate to the controllability and observability of the system. Alper (1972) discusses the meaning and relevance of these concepts in an educational planning model. Hypothesizing a simple linear dynamic system:

$$x(t+1) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t)$$

where the first equation as in previous cases relates the state of the system (number of teachers at different levels of the system) to previous states in the system and the controls (the number of graduates from teacher's colleges). The second equation is relevant in situations where the state of the system is not directly observable. Thus the outputs of the system (e.g. hours taught, salaries, etc.) is assumed to be linearly related to the state of the system. A system is said to be controllable if <u>any</u> initial state $x(o)$ can be transformed into <u>any</u> final state $x(T)$ in a finite time by some control u. A system is said to be observable if <u>every</u> state $x(o)$ can be exactly inferred from measurements of the output over a finite interval of time. In a time-invariant system with linear dynamics, these two properties can be deduced simply by performing calculations on the ranks of various combinations and powers of the matrices A, B, and C. Less is known about the more relevant concepts of relative observability and relative controllability - i.e., the <u>degree</u> to which the system can be controlled or observed.

5.6. Control Theory: A Construction Industry Example

Let us now turn to a construction industry example. For simplicity it is assumed there is no seasonal variation. Let decisions to build in each month be a sequence of independent random variables of zero mean (Gaussian or white noise), ε_t. Output of the construction industry in period t is related to the decisions as a moving-average process in which previous orders are weighted according to their age.

$$x_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \ldots \tag{1}$$

Alternatively it is possible to conceive of current output in construction as a weighted sum of previous values plus a random shock in the present period.

$$x_t = \varepsilon_t + a_1x_{t-1} + a_2x_{t-2} + \ldots \tag{2}$$

This is an autoregressive process.

These two equations do represent rather different models substantively and rather than choose between them we combine them as:

$$x_t = a_1x_{t-1} + a_2x_{t-2} + \ldots a_px_{t-p} + \varepsilon_t - b_1\varepsilon_{t-1} - b_2\varepsilon_{t-2} - \ldots - b_q\varepsilon_{t-q} \tag{3}$$

This is the mixed model of Box and Jenkins (1970) and Jenkins and Watts (1968). Taking Z transforms, equation 1 becomes

$$x_t = (1 + b_1 Z^{-1} + b_2Z^{-2} + \ldots)\ \varepsilon_t = b(Z)\ \varepsilon_t$$

and equation 2 is

$$\varepsilon_t = (1 - a_1Z^{-1} - a_2Z^{-2} - \ldots)\ x_t = a(Z)\ \hat{x}_t$$

so that equation 3 may be written as

$$x_t = \frac{b(Z)}{a(Z)}\ \varepsilon_t = \left(\frac{1-b_1\ Z^{-1} - \ldots - b_q\ Z^{-q}}{1-a_1\ Z^{-1} - \ldots - a_q\ Z^{-p}}\right)\ \varepsilon t \tag{4}$$

The term in brackets is known as the transfer function which relates the decisions to build with the output of the construction industry. It may be seen that this model has the interesting property of not implying one way causation but rather there is mutual adjustment.

In order to take account of non-stationarity, instead of the expression

$$a(Z)\, x_t = b\,(Z)\, \varepsilon_t$$

Box and Jenkins wrote

$$a\,(Z)\, \nabla^d x_t = b\,(Z)\, \varepsilon_t \tag{5}$$

∇^d is a difference operator defined as

$$a(Z)\, x_t = a(Z)\, (1-Z)^d\, x_t = a(Z)\, \nabla^d\, x_t$$

where a (Z) is stationary. With d = 1 non-stationarity in general level is allowed; with d = 2, non-stationarity in general slope is taken care of. Essentially the stationary process (d = 0) is integrated d times.

$$a(Z)\,(1-Z)^d = 1 - a_1\, Z^{-1} - a_2\, Z^{-2} - \ldots - a_{p+d}\, Z^{p+d}$$

where p is the order of the autoregression, i.e. the number of lags, and d the order of the difference.

With p = 1, d = 1, q = 1 (the order of the moving average), equation 5 becomes

$$(1 - a_1\, Z^{-1})\,(1 - Z^{-1})\, x_t = (1 - b_1\, Z^{-1})\, \varepsilon_t$$

so that

$$x_t = (1 + a_1)\, x_{t-1} - a_2\, x_{t-2} + \varepsilon_t - b_1 \varepsilon_{t-1}$$

It may happen that there is a delay factor as dead time before construction responds to a decision to build.Using a difference equation form of equation 4 we may write

$$(1 - \delta_1 Z^{-1} - \ldots - \delta_r Z^{-r})\, y_t = (1 - w_1 Z^{-1} - \ldots - w_s Z^{-s})\, x_{t-b} \qquad (6)$$

The b subscript represents the dead-time. Define

$$\Omega(Z) = w(Z)\, Z^b$$

so that the model is

$$\delta(Z)\, y_t = \Omega(Z)\, x_t$$

and the transfer function

$$V(Z) = \frac{y_t}{x_t} = \delta^{-1}(Z)\, \Omega(Z)$$

Substituting in Equation 6

$$(1 - \delta_1 Z^{-1} - \ldots - \delta_r Z^{-r})\,(V_o + V_1 Z^{-1} + V_2 Z^{-2} + \ldots) =$$
$$(1 - w_1 Z^{-1} - \ldots - w_s Z^{-s})\, Z^{-b}$$

Sorting out the coefficients

$$V_{j < b} = 0$$

$V_{b \leq j \leq b + s}$: in this range the system is settling down; all the $\delta_i V_j$'s are added in but an increasing number of w_j's are subtracted so that no pattern emerges.

$V_{j > b + s} = \delta_1 V_{j-1} + \delta_2 V_{j-2} + \ldots + \delta_r V_{j-r}$: the equation is now controlling the system.

We now wish to establish controls which will go some way towards stabilizing output of construction.

5.6.1. Feedforward Control

The first type of scheme is feedforward control: here one or more sources of disturbance may be measured and compensating deviations are introduced to minimize the mean square error of the output.

In Box and Jenkins terminology, the effect of the observed input disturbance, Z_t, is

$$\delta^{-1} (B)\ w\ (B)\ Z_{t-b}$$

Similarly, the effect of the compensation, X_t, is

$$L_1^{-1} (B)\ L_2\ (B)\ x_{t-f-1+}$$

The functions are polynominals in B and are transfer functions for a pulsed input. The effect of the observed disturbance will be cancelled if we set

$$L_1^{-1} (B)\ L_2\ (B)\ x_{t-f-1+} = -\delta^{-1} (B)\ w\ (B)\ Z_{t-b}$$

Thus, the control action at time t should be such that

$$L_1^{-1}(B)\ L_2\ (B)\ x_{t\ +} = -\delta^{-1} (B)\ w\ (B)\ Z_{t-\ (b-f-1)}$$

When $b \geq f + 1$, the control action at time t is to set x to the level

$$x_{t+} = -\frac{L_1\ (B)\ w\ (B)}{L_2\ (B)\ \ \delta(B)}\ Z_{t\ -\ (b-f-1)}$$

which will exactly balance the disturbance.

Control can also be achieved in terms of changes in Z and x.

When $b < f + 1$, the disturbance reaches the output before compensating action can become effective. Assume the disturbance is generated in the form

$$Z_{t+f+1-b} = \hat{Z}_t\ (f+1-b) + e_t\ (f+1-b)$$

When the unforecastable error

$$e_t\ (f+1-b) = a_{t+f+1-b} + \psi_1 a_{t+f-b} + \ldots + \psi_{f-b}\ a_{t+1}$$

Then

$$\delta(B)\ L_2\ (B)\ x_{t+} = -\ L_1\ (B)\ w\ (B)\ \hat{Z}_t\ (f+1-b) - L_1(B)\ e_t\ (f+1-b)$$

Optimal action is set by

$$x_{t+} = - \frac{L_1 (B) \quad w (B)}{L_2 (B) \quad \delta (B)} \hat{Z}_t \quad (f+1-b)$$

An additional component in the deviation ε_t from the target results as well as the usual noise term N_t, i.e.,

$$\varepsilon_t = N_t + \delta^{-1} (B) \; w (B) \; e_{t-f-1} \; (f+1-b)$$

5.6.2. Feedback Control

Feedforward control require both measurements of the disturbing variables and knowledge of how the disturbances affect the output. This degree of understanding would rarely be available for features of the urban system, even assuming that we could design appropriate control features. It is much more likely that we could use the "error" or predicted "error" i.e., the deviation between what is happening and what is hoped would happen, as indicating the appropriate adjustment. We shall therefore examine the principles of feedback control, leaving discussion of its practicality until later.

Without control, the deviation from target at time t,

$$N_t = \phi^{-1} (B) \quad \theta (B) \; a_t$$

where a_t is a white noise process.

We would wish to set

$$x_{t+} = - L_1 (B) \; L_2^{-1} (B) \; N_{t+f+1}$$

so that the effect of the disturbance is exactly balanced. Since we cannot know N_{t+f+1} at time t we replace it by $\hat{N}_t$ (f + 1), which is its forecast value from time t. Minimum mean square error control is then achieved by taking control action

$$x_{t+} = - L_1 (B) \; L_2^{-1} (B) \; \hat{N}_t \; (f+1)$$

The error at the output at time t will be the error in the forecast for the N_t process made for lead time f + 1.

$$\varepsilon_t = N_t - \hat{N}_{t-f-1}(f+1)$$

$$= e_{t-f-1}(f+1)$$

Writing

$$N_{t+f+1} = e_t(f+1) + \hat{N}_t(f+1)$$

and recalling that each of these terms are linear functions of the noise sequence we may amend to

$$N_{t+f+1} = L_4(B)\, a_{t+f+1} + L_3(B)\, a_t$$

$$\hat{N}_t(f+1) = \frac{L_3(B)}{L_4(B)}\, e_{t-f-1}(f+1) = \frac{L_3(B)}{L_4(B)}\, e_t$$

The control equation thus becomes

$$X_{t+} = \frac{-L_1(B)\, L_3(B)\, e_t}{L_2(B)\, L_4(B)}$$

so, in terms of the adjustment to be made at time t

$$x_t = X_{t+} - X_{t-1+}$$

$$= -\frac{L_1(B)\, L_3(B)\, (1-B)\, \varepsilon_t}{L_2(B)\, L_4(B)}$$

Note that the updating formula consists of

$$\hat{N}_t(f+1) - \hat{N}_{t-1}(f+1) = L_3(B)\, (1-B)\, a_t$$

5.7. Automatic Control of Transition Probabilities

One clear possibility for controlling a process which can be modelled as a Markov chain is to have automatic modification of the transition probabilities to allow a desired goal to be reached. In some ways this is the most natural approach to controlling a Markov chain but unfortunately hardly any work appears to have been done on it. We shall outline the only study we have found and, because an urban analogue is not immediately apparent, we shall retain the substantive content of the original. The reader may prefer to think of a network of income flows between a set of towns with connections as stated. Instead of a marked element of liquid one can have a marked dollar bill.

Complex K is formed by an array of nine vessels connected by ducts. The input is coming from an imaginary vessel numbered 0 with another at the output, 10. A generator of impulses is used to introduce these impulses repeatedly, intervals between impulses being large enough so that concentration transients practically die out.

S^* are the internal state variables measured at discrete time intervals from the instant of the application of the impulse. The desired pattern of concentration of liquid in different vessels is set by the command variables R (which could change slowly over time). The method of analysis is to evaluate the internal state vector $\vec{S}^*$ at time $t_m = m\Delta T$ by a criterion of performance expressing the degree of attainment of the desired pattern set by R. On the basis of this evaluation, new values of the acting variables A, are calculated and applied. Governing of flows is by control of the transition probabilities P_{jk} between the states of the complex. A Markov process, discrete in time and space, is used to calculate responses to the unit impulses. Before making calculations it is assumed that the

complex K "is perfectly mixed."

<u>Use of Chapman-Kolmogorov Equation to Calculate Dynamic Behaviour</u>

The states of the complex are numbered according to the vessel in which a hypothetical marked element of liquid is contained. The probability that a marked element will remain in the z^{th} vessel at time Δt

$$P_{zz} = e^{-P_z \Delta t} \tag{1}$$

where the total transition probability for state z is

$$P_z = I_z : V_z \tag{2}$$

I_z (m^3/s) being the flow from the vessel z of volume $V_z(m^3)$. The sum of input flows to a vessel is equal to the sum of output flows.

After one transition from the initial state

$$S(1) = S(0)\ [P_{jk}] \tag{3}$$

where S(0) is the initial state probability vector and $[P_{jk}]$ is the matrix of transition probabilities whose elements are determined on the basis of (1) or other knowledge.

After n transitions,

$$S(n) = S\ (n-1).\ [P_{jk}]. \tag{4}$$

Dimensionless time T is introduced by taking

$$\Delta T = \Delta t\ (I_c/V_c) \tag{5}$$

where I_c is the throughput flow and V_c is the sum of the volumes of the vessels of the complex. Relative concentrations are expressed by the relation

$$\psi_z = c_z\ (T):\ c_{io} \tag{6}$$

where $C_z(T)$ is the concentration in the vessel z at time T (after application of a unit impulse and

$$C_{io} = Q : V_{z_i}$$

is the concentration in vessel z_i, where the impulse has been applied, Q being the quantity of material injected.

From (3) and (4) we have

$$S(n) = S(0) . [P_{jk}]^n \tag{7}$$

which is the matrix form of the Chapman-Kolmogorov functional equation where $[P_{jk}]^n$ is the matrix of n-step transition probabilities.

Choosing S(0) in the form

$$S(0) = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0) \tag{8}$$

the elements of the row vector S(n) give the values at the time corresponding to n-steps of the approximations of the relative concentrations ψ_z

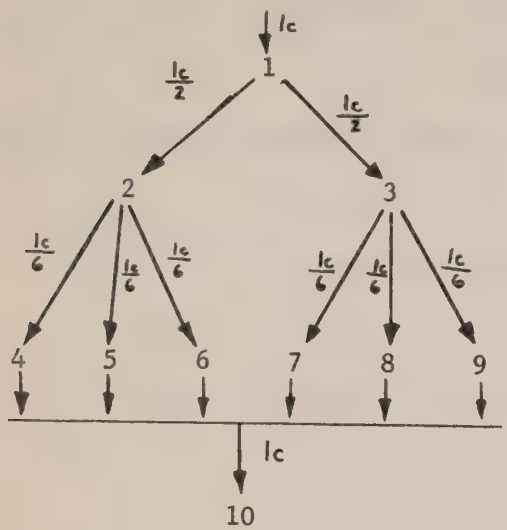
$$(z = 1, 2, \ldots\ldots, 10)$$

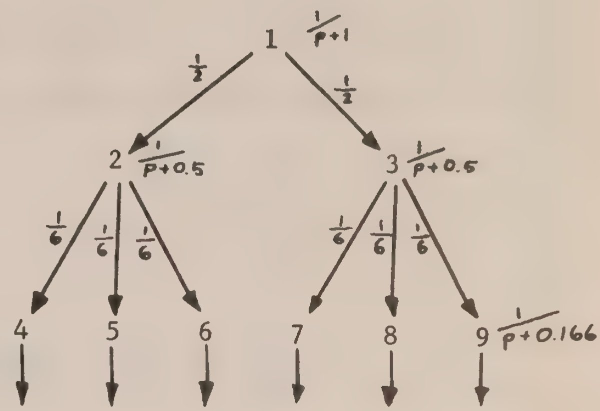
after application of a unit impulse of concentration to vessel 1.

5.7.1 The Method of the Algebra of Block Diagrams with Weighting Coefficients V_{jk}

The responses (the transients of concentration) of the different vessels of the complex to a unit impulse of concentration can be computed as the inverse Laplace transformations of the transfer functions of the respective structures built by connecting linear aperiodic elements and introducing into the algebra of block diagrams certain weighting coefficients V_{jk} which indicate the relative positions of the total throughput flow which exist in the links j - k. The transfer functions of the

different aperiodic elements are determined with a view to equations(2) and (1), i.e., in accordance with the flow I_z leaving the different vessels (and with a view to the volumes V_z of these vessels).





The volumes of the vessel are equal. The concentration transient in vessel 4 following a unit impulse of concentration at the input of the complex may be expressed as

$$\psi_4(t) = L^{-1}\left[1. \frac{1}{p+1} \cdot \frac{1}{2} \cdot \frac{1}{p+0.5} \cdot \frac{1}{6} \cdot \frac{1}{p+0.1667} \right] \tag{9}$$

Using the relation

$$L^{-1}\left[\frac{1}{(p+\alpha)\ (p+y)(p+\delta)}\right] = \frac{e^{-\alpha t}}{(y-\alpha)(\delta-\alpha)} + \frac{e^{-yt}}{(\alpha-y)(\delta-y)} + \frac{e^{-\delta t}}{(\alpha-\delta)(y-\delta)}$$

we have

$$\psi_4(t) = \frac{1}{12}\left(\frac{e^{-t}}{0.41667} + \frac{e^{-0.5t}}{0.1667} + \frac{e^{-0.1667t}}{0.277774}\right)$$

For example, for $t = 1.8$ the value of the relative concentration is

$$\psi_4(1.8) = 0.052245.$$

5.7.2. Types of Complex

1. The open-loop complex might have say nine vessels with a tenth container from which the liquid which enters it cannot return to the other nine vessels. Analysis will require an (N-1).(N-1) simple stochastic matrix of transition probabilities $[P_{jk}]$.

2. The closed-loop complex requires an N x N doubly stochastic matrix $[P_{jk}]$ i.e.

$$P_{jk} \geq 0$$

$$\sum_{k=1}^{N} P_{jk} = 1 \quad ; \quad j = 1, 2, \ldots, N$$

$$\sum_{j=1}^{N} P_{jk} = 1 \quad ; \quad j = 1, 2, \ldots, N$$

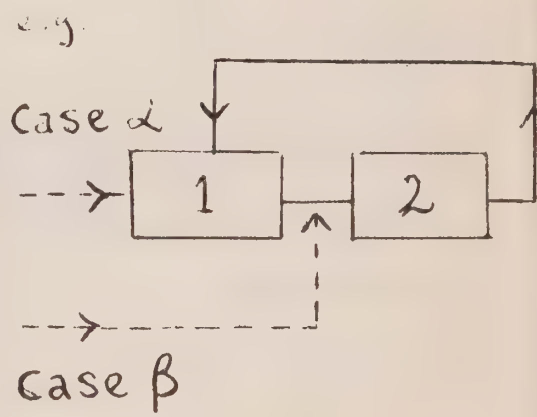
For example,

let I_c (throughput flow) = 1

V_c (total volumes) = 1

$V_1 = V_2 = V_c : 2$

$\Delta T = 0.01$



Case α

$$[P_{jk}] = \begin{pmatrix} .98020 & .01980 \\ .01980 & .98020 \end{pmatrix}$$

The initial state probability vector being

S(0) = (1, 0), we have

S(1) = .98020 .01980

S(2) = .96118 .03882, etc.

For $n \to \infty$; $S(n) \to (0.5 \; ; \; 0.5)$

Case β

$$S(0) = 0 \quad ; \quad 1$$

$$S(1) = .01980 \quad ; \quad .98020$$

$$S(2) = .03882 \quad ; \quad .96118, \text{ etc.}$$

For $n \to \infty$ $\quad S(n) \to 0.5 \quad ; \quad 0.5$

Case γ

$$[P_{jk}] = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$S(0) = 1 \quad ; \quad 0$$

$$S(1) = S(2) = \ldots\ldots = S(n) = 0.5 \quad ; \quad 0.5$$

To obtain

$$P_{11} = P_{12} = P_{22} = P_{21} = 0.5 \qquad (\Delta T = 0.01)$$

$$P_{11} = e^{-2I_c \Delta t/V_c} = e^{-0.02x} = 0.5$$

$$x = 34.6$$

$$\text{i.e.} \quad I_c' = 34.6\ I_c$$

This very large flow is required to bring the two concentrations quickly (in one step) to an equal value in both vessels.

A criterion for evaluation of the performance of a complex is required. Many are possible. The distance between the end point of the state vector $\vec{S}$ and the end point of the desired state vector $\vec{R}$ is computed in n-dimensional euclidean space. $\vec{R}$ is set by the command variables. For n=10,

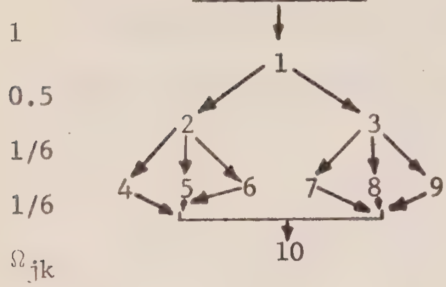
$$\delta = [(S_1 - r_1)^2 + (S_2 - r_2)^2 + (S_3 - r_3)^2 + \ldots + (S_{10} - r_{10})^2]^{0.5}$$

with

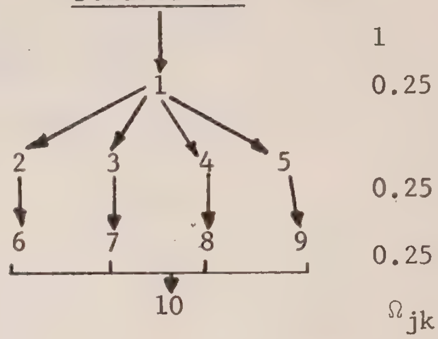
$$\sum_{z=1}^{10} S_z = 1 \quad \text{--- the unit impulse.}$$

We take $\vec{S}(t_m) - \vec{R}(t_m)$ to discuss the time to reach the desired state. If $\vec{S} - \vec{R}$ is greater than permitted, another $[P_{jk}]$ must be used.

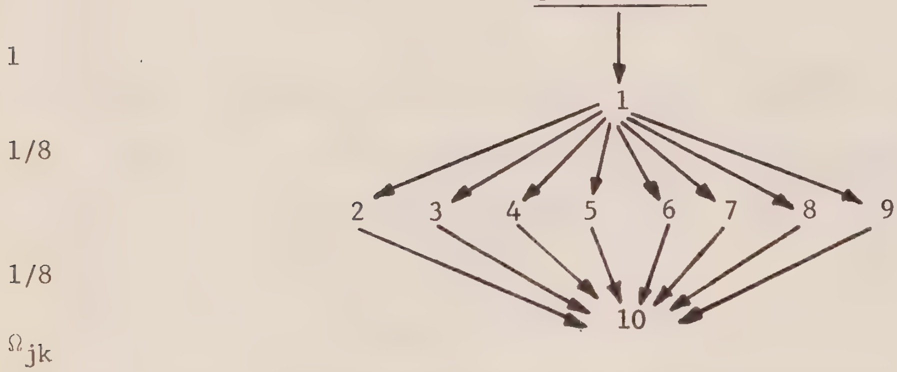
Structure A



Structure B



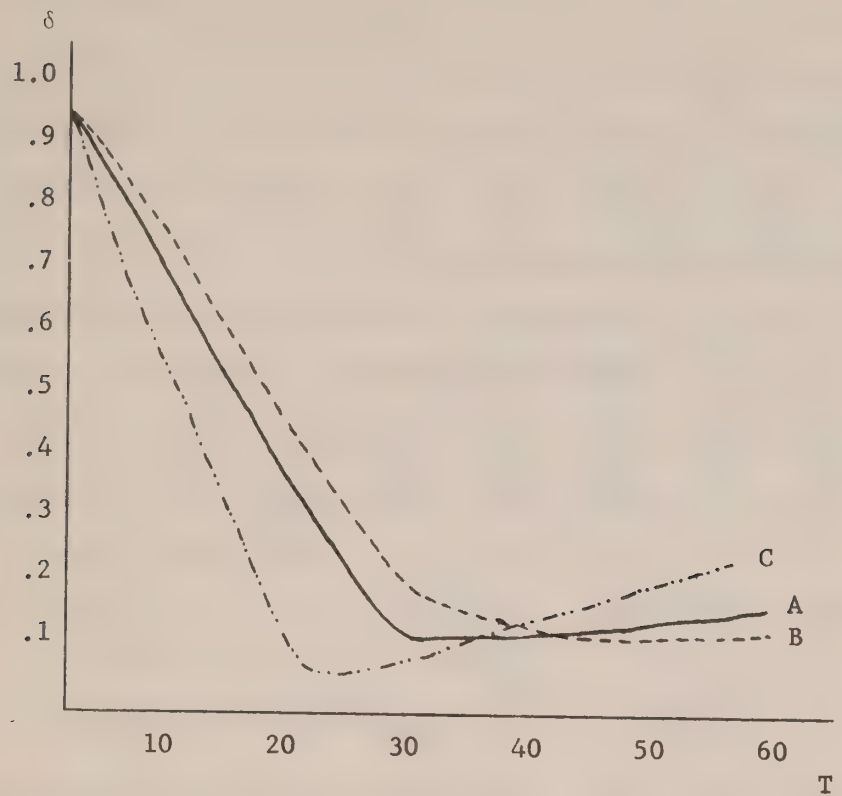
Structure C

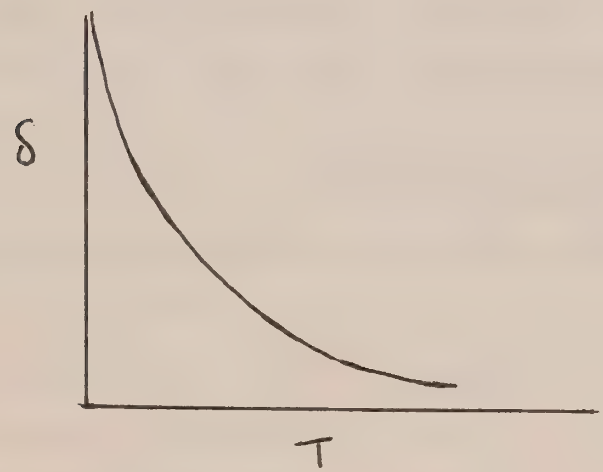


$$S(0) = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$\vec{R} = (.01, .01, .01, .01, .01, .01, .01, .01, .01, .01)$$

Results

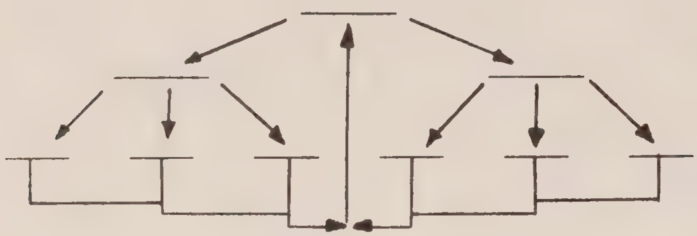




Changing Transition Probabilities

1/2

1/6



Volumes of Vessel

$$[P_{jk}] = \begin{matrix} .9139 & .0430 & .0430 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .9570 & 0 & .0143 & .0143 & .0143 & 0 & 0 & 0 \\ 0 & 0 & .9570 & 0 & 0 & 0 & .0143 & .0143 & .0143 \\ .0143 & 0 & 0 & .9857 & 0 & 0 & 0 & 0 & 0 \\ .0143 & 0 & 0 & 0 & .9857 & 0 & 0 & 0 & 0 \\ .0143 & 0 & 0 & 0 & 0 & .9857 & 0 & 0 & 0 \\ .0143 & 0 & 0 & 0 & 0 & 0 & .9857 & 0 & 0 \\ .0143 & 0 & 0 & 0 & 0 & 0 & 0 & .9857 & 0 \\ .0143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9857 \end{matrix}$$

$S(0) = (1, 0, 0, 0, 0, 0, 0, 0, 0)$

$R = (.111, .111, .111, .111, .111, .111, .111, .111, .111)$

5.8. Maximum Entropy and Land Development Processes

Consider a city council charged with the task of preparing sites for development. It must prepare these in advance of demand and keep a stock available. Obviously it wants to keep its inventories small since it costs money in reduced rates, etc., to hold cleared land, yet at the same time it endeavours to have sites available on demand, i.e., within a month, to meet uncertain orders. Assume there are three classes of land-use, say, industrial, apartments, and low-rise housing estates. From past experience, the average monthly total acreage required for each use is known; other than stocks in hand, this is all the information available.

Acres	Housing	Apartments	Industrial
In stock	100	150	50
Average Monthly demand	50	100	10

200 acres can be made ready each month and we assume that only one land-use can be handled in any one month. Intuitively, it would appear that apartment land should be readied but we provide a mathematical formalism for this simple case so that it is available for more complex cases.

We assign probability distributions to represent our state of uncertainty about possible orders in the next month for each class of land and allow the average monthly demand to represent their expectations. Allowing the entropy measure, $S = - \Sigma\, {}_{p}k \ \log_{p}k$ to represent the best measure of uncertainty contained in the probability density function to be assigned, then that function which has maximum entropy, subject to the known average value, is that which should be chosen. It is best in the sense that in maximizing uncertainty it makes the least assumptions not

warranted by the evidence. Each land use demand function comprises probabilities P_j corresponding to possible demands, Θ_j with products $n_{1,2,3}$ and expectations $< n_{1,2,3} >$. The P_j are assigned to maximize S subject to the $< n_i >$ known. Lagrange multipliers, λ_i, are required in the solution and are fixed in value by forcing agreement with the given averages. The results are summarized by the partition function.

$$Z(\lambda_1, \lambda_2, \lambda_3) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=0}^{\infty} \exp(-\lambda_1 n_1 - \lambda_2 n_2 - \lambda_3 n_3)$$

$$= \prod_{i=1}^{3} (1 - e^{-\lambda i})^{-1}$$

$$< n_i > = 1 / e^{\lambda i} - 1$$

The maximum entropy probability assignment for P_j

$$P_i(n_i) = (1 - e^{-\lambda i})\, e^{-\lambda i n_i}$$

$$= \frac{1}{<n_i> + 1} \left(\frac{< n_i >}{<n_i >+ 1} \right)^{n_i}$$

It is reasonable to assume that losses are measured by the number of acres asked for but not supplied. Let each type of land use preparation be chosen successively for the month's run and compare losses. Inventories are denoted by I_j.

1. Housing:

$$\text{Loss} = g\,(n_1 - I_1 - 200) + g\,(n_2 - I_2) + 9\,(n_3 - I_3)$$

2. Apartments:

$$\text{Loss} = g\,(n_1 - I_1) + g\,(n_2 - I_2 - 200) + g\,(n_3 - I_3)$$

3. Industrial:

$$\text{Loss} = g\,(n_1 - I_1) + g\,(n_2 - I_2) + g\,(n_3 - I_3 - 200)$$

where $$g(x) \equiv \begin{cases} x, & x \geq 0 \\ 0, & x \leq 0. \end{cases}$$

For decision 1 the expected loss will be:

$$\langle L \rangle_1 = \sum_{n_1} P_i(n_1)\, g(n_1 - I_1 - 200) + \sum_{n_2} P_2(n_2)\, g(n_2 - I_2) + \sum_{n_3} P_3(n_3)\, g(n_3 - I_3)$$

and similarly for the other decisions.

Summing, one obtains for decision 1

$$\langle L \rangle_1 = \langle n_1 \rangle e^{-\lambda_1 (I_1 + 200)} + \langle n_2 \rangle e^{-\lambda_2 I_2} + \langle n_3 \rangle e^{-\lambda_3 I_3}$$

with a similar procedure for other decisions

this gives

$\langle L \rangle_1 = 22.4$ acres of unfilled orders

$\langle L \rangle_2 = 9.7$ acres of unfilled orders

$\langle L \rangle_3 = 28.9$ acres of unfilled orders

This squares with our intuitive feeling about apartment land being needed most. However we now have a method for handling more complicated problems.

As well as knowing the average monthly demand for each land-use let us now know the average individual order for each use. Our knowledge now consists of:

Acres	Housing	Apartments	Industrial
In stock	100	150	50
Average Monthly Demand	50	100	10
Average individual order	75	10	20

Let there be a frequency distribution of orders for each type with u_h

individual orders for h acres each for housing, v_a individual orders for a acres each of apartments and w_i individual orders for i acres each for industry. The monthly demand for housing

$$n_1 = \sum_{h=1}^{\infty} hu_h$$

and the total number of individual orders for housing lots

$$m_1 = \sum_{h=1}^{\infty} u_h$$

with similar sums for apartments and industry.

As well as knowing $<n_1>$ etc., we now know $<m_1>$ etc.

With six average values given, six Lagrange multipliers are necessary:

$$\lambda_1, \mu_1, \lambda_2, \mu_2, \lambda_3, \mu_3.$$

The partition function factors with each factor having the form

$$Z_1(\lambda_1 \mu_1) = \sum_{u_1=1}^{\infty} \sum_{u_2=1}^{\infty} \ldots \exp(-\lambda_1 (u_1 + 2u_2 + 3u_3 + \ldots)$$
$$-\mu_1 (u_1 + u_2 + u_3 + \ldots))$$

$$= \prod_{h=1}^{\infty} \frac{1}{1-e^{h\lambda_1 - \mu_1}}$$

We then have

$$<n_1> = \sum_{h=1}^{\infty} \frac{h}{e^{h\lambda_1 + \mu_1} - 1} \simeq \sum_{n=1}^{\infty} \frac{e^{-n(\mu_1 + \lambda_1)}}{(1 - e^{-n\lambda})^2}$$

$$<m_1> = \sum_{h=1}^{\infty} \frac{1}{e^{h\lambda_1 + \mu_1} - 1} \simeq \sum_{n=1}^{\infty} \frac{e^{-n(\mu_1 + \lambda_1)}}{1 - e^{n\lambda}}$$

so that

$$<u_h> = \frac{1}{e^{h\lambda_1 + \mu_1} - 1} \tag{1}$$

The second expressions for $< n_1>$ and $< m_1>$ are approximate transforms which allow the series represented in the first expressions to converge rapidly. Using these,

$$< n_1 > \;\simeq\; \frac{e^{-\mu_1}}{\lambda_1^{2}} \qquad (2)$$

$$< m_1 > \;\simeq\; \frac{e^{-\mu_1}}{\lambda_1}$$

$$\lambda_1 \;\simeq\; \frac{< m_1>}{< n_1>} = \frac{1}{75}$$

$$e^{\mu_1} \;\simeq\; \frac{< n_1>}{< m_1>^2} = 112 \;; \qquad \mu_1 = 4.72$$

The probability that u_h has a particular value h,

$$P(u_h) = (1 - e^{h\lambda_1 - \mu_1}) \, e^{-(h\lambda_1 + \mu_1)u_h}$$

with mean value (1) and variance

$$\mathrm{Var}\,(u_h) = < u_h^{\,2} > - < u_h>^2 = \frac{e^{h\lambda_1 + \mu_1}}{(e^{h\lambda_1 + \mu_1} - 1)^2}$$

The probability distribution of total demand for housing $P(n_1)$ will be very nearly Gaussian because it represents the sum of a large number of independent random variables, hu_h. Its mean value is as (2) and variance

$$\mathrm{Var}\,(n_1) = \sum_{h=1}^{\infty} h^2 \,\mathrm{Var}\,(u_h)$$

$$= \sum_{h=1}^{\infty} \frac{h^2 \, e^{h\lambda_1 + \mu_1}}{(e^{h\lambda_1 + \mu_1} - 1)^2}$$

$$\simeq \frac{2e^{-\mu_1}}{\lambda_1^{\,3}} = \frac{2}{\lambda_1} < n_1 >$$

We thus obtain

$$P(n_1) \simeq \left(\frac{\lambda 1}{4\Pi < n_1 >}\right)^{0.5} \exp\left(- \frac{\lambda_1 (n_1 - <n_1>)^2}{4 < n_1 >}\right)$$

This equation will provide a good estimate for large values of n_1 but a poor approximation for n_1 small.

Expected loss consequent on a decision is as before, the sum of the three terms for failure to meet orders for each type of land use. Taking separately the loss on housing if no housing sites are prepared,

$$\sum_{n_1=0}^{\infty} P(n_1)\, g(n_1 - I_1) \simeq \left(\frac{\lambda_1}{4\,\Pi < n_1 >}\right)^{0.5} \int_{I_1}^{\infty} (n_1 - I_1) \exp\left(- \frac{\lambda_1 (n_1 - < n_1 >)^2}{4 < n_1 >}\right) dn_1$$

$$\sim \frac{< n_1 > - I_1}{2} \left(1 + \operatorname{erf} \alpha_1 \left(< n_1 > - I_1\right)\right) + \frac{1}{2\alpha_1 \Pi^{0.5}} \exp\left(- \alpha_1^2 \left(< n_1 > - I_1\right)^2\right) \qquad (3)$$

where $\alpha_1^2 = \lambda_1 / 4 < n_1 >$ and erf χ is the error function

$$\operatorname{erf} \chi = \frac{2}{\Pi^{0.5}} \int_0^{\chi} e^{-\chi^2}\, d\chi$$

If housing sites are produced, I_1 is replaced by $(I_1 + 200)$ in (3). Residential estates and industry are handled similarly. Making the necessary calculations with $\alpha_1 = 0.0082$, $\alpha_2 = 0.016$ and $\alpha_3 = 0.035$, we find

$< L >_1 = (0) + 2.86 + 0.18 = 3.04$ acres of unfilled demand

$< L >_2 = 14.9 + (0) + 0.18 = 15.10$ acres of unfilled demand

$< L >_3 = 14.9 + 2.86 + (0) = 17.80$ acres of unfilled demand

with (0) being a negligibly small term.

Whereas previously we chose apartment land on the basis of average monthly demand, we now choose residential estates with the additional information about average individual sites required.

We may keep on refining our choice as further information is supplied. Suppose for example that we learn that a specific order for a 40 acre site for industry must be met in the following month. The same equations apply as in the previous example save that I_3 is reduced from 50 to 10. Then $(<n_3> - I_3) = 0$ so that (3) reduces to $1/2\, \alpha_3\, \Pi^{0.5} = 8.08$. Losses now are

$$< L >_1 = (0) + 2.86 + 8.08 = 10.94 \text{ acres}$$

$$< L >_2 = 14.9 + (0) + 8.08 = 23.00 \text{ acres}$$

$$< L >_3 = 14.9 + 2.86 + (0) = 17.80 \text{ acres}$$

Thus, the new order does not change the decision to concentrate on housing estate sites although this was not obvious intuitively.

Although a land use example has been chosen for illustration, this methodology should find many policy applications where decisions are to be made about allocating productive capacity to various ends when only a minimum amount of information is available.

APPENDIX A

List of References

Adelman, I.G. 1958. A stochastic analysis of the size distribution of firms. Journal of the American Statistical Association 53:893-904.

Alper, P. 1972. Some considerations of planning engendered by modern control theory. Socio-Economic Planning Sciences 6:295-303.

Anderson, T.W. and L. Goodman. 1957. Statistical inference about Markov chains. Annals of Mathematical Statistics 28:89-110.

Arrow, K.J., S. Karlin, and H. Scarf. 1958. Studies in the Mathematical Theory of Inventory and Production. Stanford: Stanford University Press.

Athans, M. and P.L. Falb. 1966. Optimal Control: An Introduction to the Theory and Its Applications. New York: McGraw-Hill.

Bailey, F.N. and R.T. Holt. 1971. The Analysis of Governmental Structures in Urban Systems. I.E.E.E., Systems, Man and Cybernetics Group, Annual Symposium Record. Anaheim, California.

Barr, J.L. 1972. City size, land rent, and the supply of public goods. Regional and Urban Economics 2:67-103.

Bartholomew, D.J. 1967. Stochastic Models for Social Processes. New York: Wiley.

Batty, M. 1971. Modelling cities as dynamic systems. Nature 231:425-428.

_______. 197 . An experimental model of urban dynamics. Journal of Town and Country Planning

_______. 1973. A probability model of the housing market based on quasi-classifical considerations. Socio-Economic Planning Sciences 7:6:573-598.

Barber, G. 1970. Dynamic locational analysis: the firm and its operating policy. Discussion Paper No. 9. Department of Geography, University of Toronto.

Baumol, W.J. 1972. The dynamics of urban problems and its policy implications. In M. Preston and B. Corry, eds. Essays in Honour of Lord Robbins. London: Weidenfeld and Nicholson, pp. 380-393.

Bebee, E. 1972. Regional housing markets and population flows in Canada: 1956-1967. Canadian Journal of Economics 5:3:386-397.

Bell, F.W. 1967. An econometric forecasting model for a region. Journal of Regional Science 7:109-128.

Bharucha-Reid, A.T. 1960. Elements of the Theory of Markov Processes and their Applications. New York: McGraw-Hill.

Berstein, S.J., et al. 1972. Urban renewal: an investment policy analysis employing linear programming. Socio-economic Planning Sciences 6:3: 251-259.

Benes, J. 1970. Pattern evolution through automatic control of transition probabilities. In A.S. Iberall and J.B. Reswick, eds. Technical and Biological Problems of Control - A Cybernetic View. Pittsburg, Penn.: Int. Fed. of Automatic Control. pp. 817-823.

Berry, B.J. L. 1971. Monitoring trends, forecasting change and evaluating goal-achievement in the urban environment: the ghetto expansion vs. disegregation issue in Chicago as a case study. In M. Chisholm, A.E. Frey and P. Haggett, eds. Regional Forecasting. London: Butterworths. pp. 93-117.

Bhat, U.N. 1972. Elements of Applied Stochastic Processes. New York: Wiley.

Blumberg, D.F. 1971. The city as a system. Simulation 17:155-167.

Blumen, I., M. Kogan and P.J. McCarthy. 1955. The Industrial Mobility of Labour as a Probability Process. Ithaca: Cornell University Press.

Borcherding, D.R. and R.T. Deacon. 1972. The demand for the services of non-federal governments: an econometric approach to collective choice. American Economic Review 62:891-901.

Borts, G.H. 1968. Growth and capital movements among U.S. regions in the postwar period. American Economic Review 58:155-161.

Bourne, L.S. 1969. Forecasting land occupancy changes through Markovian probability matrices: a central city example. Research Paper No. 14. Toronto: Centre for Urban and Community Studies, University of Toronto.

_______. 1969. A spatial land use conversion model of urban growth. Journal of Regional Science 9:261-272.

Bourne, L.S. and C.A. Maher. 1970. Comparative extrapolations of city-size distributions: the Ontario-Quebec urban system. Research Paper No. 27. Toronto: Centre for Urban and Community Studies, University of Toronto.

Bourne, L.S. and R. D. MacKinnon. 1972. Urban System Development in Central Canada: Selected Papers. Toronto: University of Toronto Press.

Bourne, L.S., R.D. MacKinnon and J.W. Simmons. 1973. The Form of Cities in Central Canada: Selected Papers. Toronto: University of Toronto Press.

Bourne, L.S., R.D. MacKinnon, J. Siegel, and J.W. Simmons. 1974. Urban Futures for Central Canada: Perspectives on Forecasting Urban Growth and Form. Toronto: University of Toronto Press.

Box, G.E.P. and G.M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. San Francisco: Holden Day.

Bradford, D.F. and H.H. Kelejian. 1973. An Econometric model of the flight to the suburbs. Journal of Political Economy 81:3:566-589.

Brown, L.A. 1970. On the use of Markov chains in movement research. Economic Geography 46:393-403.

Brown, L.A. and F.E. Horton. 1970. Functional distance: an operational approach. Geographical Analysis 2:76-82.

Brown, H.J., et al. 1972. Empirical Models of Urban Land Use: Suggestions on Research Objectives and Organization. New York: National Bureau of Economic Research.

Bryson, A.E., Jr. and Y. Ho. 1969. Applied Optimal Control. New York: Blaisdell.

Burford, R.L. 1973. A queuing theoretic interpretation of zero population growth. Baton Rouge, Louisiana: Louisiana State University. Mimeographed.

Cassidy, R.G., M.J. Kirby and W.M. Raike. 1971. Efficient distribution of resources through two or more levels of government. Management Science 17:8:462-473.

Casetti, E. 1969. On the sensitivity of regional incomes to export fluctuations. Discussion Paper No. 2. Department of Geography, The Ohio State University.

Centre for Real Estate and Urban Economics. 1968. Jobs, People and Land. Special Report No. 6. University of California, Berkeley, California.

Chapin, Francis S. 1965. A Model for Simulating Residential Development. Journal, American Institute of Planners 43:2:120-125.

Chapin, Francis S. and Shirley F. Weiss. 1962. Factors Influencing Land Development. Chapel Hill: Institute for Research in Social Science.

Chapin, Francis S. and Shirley F. Weiss. 1965. Some Input Refinements for a Residential Model. Chapel Hill: Institute for Research in Social Science.

Charnes, A., et al. 1973. A model to study revenue sharing and account for regionalized economic activity and social goals. Management Science 19:10:1189-1208.

Clark, W.A.V. 1965. Markov chain analysis in geography: an application to the movement of rental housing areas. Annals, Association of American Geographers 55:351-359.

Cohen, G. n.d. Dynamic programming. Occasional Paper No. 9. Centre for Urban and Regional Research, University of Manchester.

Cohen, J.E. 1972. Markov population processes as models of primate social and population dynamics. Theoretical Population Biology 3:119-134.

Chinitz, B. 1971. National Policy for regional development. In J.F. Kain and J.R. Meyer, eds. Essays in Regional Economics. Cambridge: Harvard University Press, pp. 21-37.

Colenutt, R.J. 1969. Linear diffusion in an urban setting: An example. Geographical Analysis 1:106-114.

Collins, L. 1970. Markov chains and industrial migration: forecasting aspects of industrial activity in Ontario. Unpublished Ph.D. thesis. Department of Geography, University of Toronto.

Cooper, W.W., N. Johnson, and K.O. Kortanek. 1971. Systems approach to urban planning: Mixed, conditional, adaptive and other alternatives. Policy Sciences 1:4:397-405.

Cordey-Hayes, M. 1972. Dynamic frameworks for spatial models. Socio-Economic Planning Sciences 6:365-385.

Cordey-Hayes, M. and D. Gleave. 1973. Migration movements and the differential growth of city regions in England and Wales. Centre for Environmental Studies Research Paper No. 1. London: CES.

Cowan, P., et al. 1967. Approaches to urban model building. Regional Studies 1:2:163-172.

Courtney, J.F., Jr., T.D. Klastorin, and T.W. Ruefli. 1972. A goal programming approach to urban - suburban location preferences. Management Science 18:6:258-268.

Cox, D.R. and H.D. Miller. 1965. The Theory of Stochastic Processes. London: Methuen

Crecine, J.P. 1964. A Time-Oriented Metropolitan Model for Spatial Location. Pittsburgh, Penn.: Department of City Planning

Crecine, J.P. 1968. A Dynamic Model of Urban Structure. Santa Monica, California: RAND Corporation.

Crecine, J.P. 1967. Computer Simulation in Urban Research. Santa Monica, California: RAND Corporation.

Crecine, J.P. 1969. Spatial Location Decisions and Urban Structure: A Time-Oriented Model. Ann Arbor: Institute for Public Policy Studies, University of Michigan.

Creighton, Hamburg, Inc. 1969. Transportation and Land Development - A Third Generation Model.

Curry, L. and E.B. MacDougall. 1971. Statistical spatial analysis and remotely sensed imagery. Unpublished manuscript. Department of Geography, University of Toronto.

Czamanski, S. 1964. A model of urban growth. Papers of the Regional Science Association 13:177-200.

Czamanski, S. 1965. A method of forecasting metropolitan growth by means of distributed lags analysis. Journal of Regional Science 6:35-49.

Czamanski, S. 1969. Regional econometric models: a case study of Nova Scotia. In A.J. Scott, ed. Studies in Regional Science, pp. 143-180.

Day, R.H. 1963. Recursive Programming and Production Response. Amsterdam: North-Holland.

Dagenais, M. 1973. Un modele annuel de provision pour l'economie du-Quebec. Canadian Journal of Economics 6:1:62-78.

DeCani, J.S. 1961. On the construction of stochastic models of population growth and migration. Journal of Regional Science 3:1-13.

DeLeeuw, F. 1972. The Distribution of Housing Services: First Empirical Steps. Working Paper 208-4. Washington, D.C. The urban Institute.

DeLeeuw, F. and N.F. Ekanem. 1973. Time lags in the rental housing market. Urban Studies 10:39-68.

DeLevę,G., et al. 1970. Generalized Markovian Decision Processes. Amsterdam: Mathematical Centre.

Denton, F.T. 1973. A simulation model of month-to-month labour force movement in Canada. International Economic Review 14:293-311.

Derman, C. 1962. On sequential decisions and Markov chains. Management Science 9:16-24.

Derman , C. 1970. Finite Markov Decision Processes. New York: Academic Press.

Dobell, A.R. 1969. Some characteristic features of optimal control problems in economic theory. IEEE Transactions on Automatic Control AC-14:1:39-48.

Donelly, T.G., F.S. Chapin and S.F. Weiss. 1964. A Probabilistic Model for Urban Growth. Chapel Hill: Institute for Research in Social Science.

Dorfman, R. 1969. An economic interpretation of optimal control theory. American Economic Review 59:5:817-831.

Drewett, J.R. 1969. A stochastic model of the land conversion process: An interim report. Regional Studies 3:3:269-280.

Eastman, C. and K. Kortanek. 1970. Modelling school facility requirements in new communities. Management Science 16:12:784-799.

Engle, R.F., et al. 1972. An econometric simulation model of intra-urban housing location: Housing, business, transportation, and local government. American Economic Review 62:2:87-97.

Erlenkotter, D. 1967. Two producing areas: dynamic programming solutions. In A.S. Manne, ed. Investments for Capacity Expansion, Size, Location and Time Phasing. Cambridge, Mass.: M.I.T. Press.

Erlenkotter, D. 1972. Economic Integration and dynamic location planning. Swedish Journal of Economics 74:8-18.

Fano, P.L. 1968. Organization, city size, distributions and central places. Papers, Regional Science Association 22:29-38.

Fisher, J.C. and B.R. Lawson. 1972. Spatial planning in Yugoslavia - an application of Markov chain analysis to changing settlement patterns. The Review of Regional Studies 2:195-214.

Ford, L.R. and W.H. Jago. An urban transportation simulation. Presented at O.R.S.A. National Meeting, San Francisco.

Forrester, J.W. 1969. Urban Dynamics. Cambridge, Mass.: MIT Press.

Forrester, J.W. 1969. Rebuttal to a review of J.F. Kain. Fortune 80:7:191-192.

Forrester, J.W. 1970. System analysis as a tool for urban planning. IEEE Trans. on Systems, Science and Cybernetics SSC-6:258-265.

Forrester, J.W. 1971. Counter-intuitive behaviour of social systems. Technological Review 73:52-66.

Fuguitt, G.V. 1965. The growth and decline of small towns as a probability process. American Sociological Review 30:403-411.

Gale, S. 1972. Some formal properties of Hagerstrand's model of spatial interaction. Journal of Regional Science 12:199-217.

Gale, S. 1972. Stochastic stationarity and the analysis of geographic mobility. Proceedings of International Geographical Union Conference, Montreal.

Gani, J. and D. Jerwood. 1971. Markov chain methods in chain binomial epidemic models. Biometrics 27:591-603.

Garn, H.A. 1970. An urban systems model: a critique of urban dynamics. Working Paper 113-125. Washington, D.C.: The Urban Institute.

Garrison, William L. 1962. Toward simulation models of urban growth and development. In Norborg, K. ed. The I.G.U Symposium on Urban Geography. Lund Stud. in Geog. Series B 24:91-108. Lund, Sweden.

Gates, G.R. and M. Anvari. 1970. A water balance control theoretic model for regional management. Geographical Analysis 2:19-29.

Ghali, M. 1973. Exports, investment, and regional growth. Southern Economic Journal 39:289-296.

Gilbert, G. 1972. Two Markov models of neighbourhood housing turnover. Environment and Planning 4:133-146.

Gilbert, G. 1973. Semi-Markov processes and mobility: A note. Journal of Mathematical Sociology 3:1:139-145.

Ginsberg, R.B. 1971. Semi-Markov processes and mobility. Journal of Mathematical Sociology 1:233-262.

Ginsberg, R.B. 1972a. Critique of probabilistic models: application of the semi-Markov model to migration. Journal of Mathematical Sociology 2:63-82.

Ginsberg, R.B. 1972b. Incorporating causal structure and exogenous information with probabilistic models: with special reference to choice, gravity, migration and Markov chains. Journal of Mathematical Sociology 2:83-103.

Ginsberg, R.B. 1973. Stochastic models of residential and geographic mobility for heterogeneous populations. Environment and Planning 5:113-124.

Glickman, N. 1971. An econometric forecasting model for the Philadelphia region. Journal of Regional Science 11:15-32.

Goldner, W. 1968. Projective land use model: A model for the spatial allocation of activities and land uses in a metropolitan region. B.A.T.S.C. Technical Report 219. California: Bay Area Transportation Studies Commission.

Golladay, F.L. and A.D. Sandalov. 1972. Optimal development policy in an open regional economy: A programming analysis. Journal of Regional Science 12:185-198.

Goodman, L.A. 1961. Statistical methods for the mover-stayer model. American Statistical Association Journal 56:

Goreux, L.H. and A.S. Manne, eds. 1973. Multi-Level Planning: Case Studies in Mexico. New York: American Elsevier.

Graham, A.K. 1972. Modelling city-suburb interactions. IEEE Transactions on Systems, Man and Cybernetics SMC-2:2_156-158.

Graybeal, R.S. 1966. A Simulation Model of Residential Development. Berkeley, California: University of California.

Hajnal, J. 1956. The ergodic properties of non-homogeneous finite Markov chains. Proceedings of the Cambridge Philosophical Society 52:67-77.

Harary, F., et al. 1970. A matrix approach to nonstationary chains. Operations Research 18:1168-1181.

Harris, B. 1964. A Model of Locational Equilibrium for Retail Trade. Institute for Environmental Studies, University of Pennsylvania.

Harris, B. 1969. Report on Household Projection Model. Institute for Environmental Studies, University of Pennsylvania.

Harris, C.C., Jr. 1968. A stochastic process model of residential development. Journal of Regional Science 8:29-39.

Harrison, D. and J.F. Kain. 1970. An historical model of urban form. Paper prepared for the CUE Research Conference, Chicago, Illinois, September 11-12, 1970.

Harvey, D.M. 1967. Models in the evolution of spatial patterns in human geography. In R.J. Chorley and P. Haggett, eds. Models in Geography. London: Methuen, pp.549-608.

Helly, W. 1969. A residential land use prediction model. In H. Feshbach and K.U. Ingard, eds. In Honor of Philip M. Morse. Cambridge, Mass.: MIT Press, pp. 255-265.

Hemmens, G.C. 1966. The structure of urban activity linkages. Centre for Urban and Regional Studies, Institute for Research in Social Science, University of North Carolina, Chapel Hill.

Henry, N.W., R. McGinnis, H.W. Tegtmeyer. 1971. A finite model of mobility. Journal of Mathematical Sociology 1:107-118.

Henry, N.W. 1971. The retention model: A Markov chain with variable transition probabilities. Journal of American Statistical Association 66:264-267.

Hester, J., Jr. 1970. Dispersal, segregation and technological change: a computer simulation model of the development of large metropolitan areas in the United States during the 20th Century. Unpublished Ph.D. thesis, M.I.T., Cambridge, Mass.

Hill, D.M. 1965. A growth allocation model for the Boston region. Journal American Institute of Planners 31:111-120.

Hillier, F.S. and G.J. Lieberman. 1967. Introduction to Operations Research. San Francisco: Holden-Day.

Horton, F.E. and J. Morris. 1970. A review of urban dynamics. Journal of Regional Science 10:421-423.

Howard, R.A. 1960. Dynamic Programming and Markov Processes. Cambridge, Mass.: M.I.T. Press.

Howard, R.A. 1971. Dynamic Probabilistic Systems. 2 vols; New York: Wiley.

Hudson, J. 1970. Elementary models for population growth and distribution analysis. Demography 7:3:361-368.

Ingram, G.K., J. F. Kain, and J.R. Ginn. 1972. The Detroit Prototype of the NBER Urban Simulation Model. New York: National Bureau of Economic Research - Columbia University Press.

Irwin, N.A. 1963. Review of existing land use forecasting techniques. Highway Review Board Record 88:194-199.

Jaynes, E.T. 1963. New engineering applications of information theory. In J.L. Bogdanoff and F. Kozin. Proc. First Symposium of Engineering Applications of Random Function Theory. New York: Wiley, pp. 163-203.

Jenkins, G.M. and D.G. Watts. 1968. Spectral Analysis and its Applications. San Francisco: Holden-Day.

Jewell, W.S. 1963. Markov renewal programming. Operations Research 11:938-971.

Jones, C. 1971. Markov chains and their uses in urban and regional research. Occasional Paper No. 10. Centre for Urban and Regional Research, University of Manchester.

Jutila, S.K. 1971. A linear model for agglomeration, diffusion, and growth of regional economic activity. Regional Science Perspectives 1:83-108.

Jutila, S.K. 1973. Spatial macroeconomic development (SMED). Papers, Regional Science Association 30:39-57.

Kain, J.F. 1964. The development of urban transportation models. Papers and proceedings of the Regional Science Association 14:147-173.

Kain, J.F. 1967. Postwar metropolitan development: housing preferences and auto ownership. American Economic Review 57:223-234.

Kain, J.F. and J.R. Meyer. 1968. Computer simulations, physio-economic systems, and intraregional models. American Economic Review 58:171-181.

Kadanoff, L.P. 1972. From simulation model to public policy. American Scientist 60:74-79.

Kadanoff, L.P. and H. Weinblatt. 1972. Public policy conclusions from urban growth models. IEEE Transactions on Systems, Man and Cybernetics SMC-2:159-166.

Kelley, A.C. and L.W. Weiss. 1969. Markov processes and economic analysis: The case of migration. Econometrica 37:2:280-297.

Kemeny, J.G. and J.L. Snell. 1960. Finite Markov Chains. Princeton: Van Nostrand.

Kendrick, D. 1971. Mathematical models for regional planning. Regional and Urban Economics 1:247-287.

Keyfitz, N. 1968. Introduction to the Mathematics of Population. Reading, Mass.: Addison-Wesley.

Keyfitz, N. 1972. On future population. Journal of the American Statistical Association 67:347-363.

Kibel, B.M. 1970. The Evolution of a Dynamic Planning Model. Unpublished Ph.D. thesis, University of California, Berkeley.

Kilbridge, M.D., et al. 1969. A conceptual framework for urban planning models. Management Science 15:246-266.

Koenker, R. 1972. An empirical note on the elasticity of substitution between land and capital in a monocentric housing market. Journal of Regional Science 12:299-305.

Krieger, M.H. 1971. Modelling urban change. Socio-Economic Planning Sciences 5:41-55.

Krisbergh, H. and C.W. Nelson. 1972. A search procedure for policy oriented simulations: applications to urban dynamics.

Lachene, R. 1965. Networks and the location of economic activities. Papers, Regional Science Association 14:183-196.

Lamb, D.R. 1967. Research of Existing Land Use Models, No. 1045. Pittsburgh: S.W. Pennsylvania Regional Planning Commission.

Land, K.C. 1969. Duration of residence and prospective migration: further evidence. Demography 6:133-140.

Larouche Pierre. 1965. The simulation of residential land use growth in the Montreal Region. Unpublished Master of City Planning Thesis, Yale University, New Haven, Connecticut.

Lathrop, G.T. and J.R. Hamburg. 1965. An opportunity-accessibility model for allocating growth. Journal of the American Institute of Planners 31:95-103.

Lea, A.C. 1973. Location-allocation models: A review. Discussion Paper No. 13. Department of Geography, University of Toronto.

Lee, S.M. and W.R. Sevebeck. 1971. An aggregative model for municipal economic planning. Policy Sciences 2:99-115.

Lee, T.C., G.G. Judge and T. Takayama. 1955. On estimating the transitional probabilities of a Markov process. Journal of Farm Economics 47:742-762.

Leslie, P.H. 1945. On the use of matrices in certain population mathematics. Biometrika 33:183-212.

Lewis, W.C. and J.R. Prescott. 1972. Urban regional development and growth centres: an econometric study. Journal of Regional Science 12:57-70.

Lindsay, I. and B.M. Barr. 1972. Two stochastic approaches to migration: comparison of Monte Carlo simulation and Markov chain models. Geografiska Annaler 54B:56-57.

Lithwick, N.H. 1970. Urban Canada: Problems and Prospects. Ottawa: Central Mortgage and Housing.

Arthur D. Little, Inc. 1964. San Francisco Community Renewal Program - Simulation Model for Renewal Programming. Technical Paper No. 1

MacKinnon, R.D. 1970. Dynamic programming and geographic systems. Economic Geography 46:350-366.

Madansky, A. 1959. Least squares estimation in finite Markov processes. Psychometrika 24:2

Malm, R., G. Olsson and O. Warneryd. 1966. Approaches to simulations of urban growth. Geografiska Annaler 48A:9-22.

Manne, A.S. 1961. Capacity expansion and probabilistic growth. Econometrica.

Manne, A., ed. 1967. Investments for Capacity Expansion: Size, Location and Time-Phasing. Cambridge, Mass.: M.I.T. Press.

Marble, D.F. 1964. A simple Markovian model of trip structures in a metropolitan region. Papers of the Regional Science Association, Western Section 1:150-156.

Marble, D.F. 1964. Two computer programs for the analysis of simple Markov chains. Research Report No. 6. Department of Geography, Northwestern University, Evanston, Illinois.

May, S.J. 1964. Short-term forecasting of non-residential construction. Canadian Political Science Association Conferences on Statistics, 1962 and 1963. Toronto: University of Toronto Press.

Massie, R.W. 1969. A system of linked models for forecasting urban residential growth. Master of Regional Planning Thesis, University of North Carolina.

Matras, J. 1967. Social mobility and social structure: some insights from the linear model. American Sociological Review 32: 608-614.

McFarland, D.D. 1970. Intragenerational social mobility as a Markov process: including a time-stationary Markovian model that explains observed declines in mobility rates over time. American Sociological Review 35:463-476.

McGinnis, R. 1968. A stochastic model of social mobility. American Sociological Review 33:712-722.

McDonald, G.T. 1972. Forecasting and the use of agricultural land in Ontario. Unpublished Ph.D. thesis. University of Toronto.

McLoughlin, J.R. 1969. Simulation for beginners: the planting of a sub-regional model system. Regional Studies 3:3:313-324.

McLoughlin, J.B. 1969. Leicester and Leicestershire Sub-Regional Planning Study. Leicester

McLoughlin, J.B. 1973. Control and Urban Planning. New York: Oxford University Press.

Mesarovic, M.D., D. Macko, and Y. Takahara. 1971. Theory of Hierarchical Multilevel Systems. New York: Academic Press.

Mesarovic, M.D. and A. Reisman, eds. 1972. Systems Approach and the City. Amsterdam: North Holland Pub. Co.

Mills, E. 1972. Studies in the structure of the Urban Economy. Baltimore: John Hopkins.

Miron, J. 1973a. Urban systems and economic growth. Unpublished, mimeographed paper. Department of Geography, University of Toronto.

Miron, J. 1973b. Four simple, dynamic models of urban growth. Unpublished, mimeographed paper. Department of Geography, University of Toronto.

Moody, H.T. 1970. Urban dynamics: a review of Forrester's model of an urban area. Economic Geography 46:620-625.

Morrill, R.L. 1962. Simulation of central place patterns over time. In K. Norborg, ed. The I.G.U. Symposium on Urban Geography Lund Studies in Geography, Series B, No. 24, Lund, Sweden. pp. 109-120.

Morrill, R.L. 1963. The development and spatial distribution of towns in Sweden. Annals, Association of American Geographers 53:1-14.

Morrill, R.L. 1965. Expansion of the urban fringe, a simulation experiment. Papers and Proceedings, Regional Science Association 15:185-199.

Morrill, R.L. 1965. The Negro Ghetto: problems and alternatives. Geographical Review 55: 339-361.

Morrill, R.L. 1965. Migration and the spread and growth of urban settlement. Lund Studies in Geography, No. 26B. Gleerup, Lund, Sweden.

Morrison, P.A. 1967. Duration of residence and prospective migration: a stochastic model. Demography 4: 553-561.

Morrison, P.A. 1973. Theoretical issues in the design of population mobility models. Environment and Planning 5:125-134.

Moses, L. and H.F. Williamson. 1967. The location of economic activities in cities. American Economic Review 57:211-222.

Muth, R.F. 1971. The derived demand for urban residential land. Urban Studies 8

Muth, R.F. 1971. Migration; chicken or egg? Southern Economic Journal 37:295-306.

Myers, G. C., R. McGinnis, and G. Masnick. 1967. The duration of residence and prospective migration. Demography 4:553-561.

Myers, G.C., R. McGinnis and G. Masnick. 1967. The duration of residence approach to a dynamic stochastic model of internal migration. Eugenics Quarterly 14:

Nagelberg, M. 1970. Simulation of Urban Systems - a Selected Bibliography. Institute for the Future, Middletown, Connecticut.

Nagelberg, M. 1970. Selected Urban Simulations and Games. Institute for the Future, Middletown, Connecticut.

Nathanson, J. 1970. A Model for the Intra-County Location of Basic Employment and Land. BATSC Tech. Report 222. Bay Area Transportation Study Commission, Berkeley, California.

Newling, B.E. 1970. Review of urban dynamics. Geographic Review 60:455-457.

Niedercorn, J.H.1963. An econometric model of metropolitan employment and growth. RM-3758-RC. Rand Corporation.

Niedercorn, J.H. 1971. A negative exponential model of urban landuse densities and its implications for metropolitan development. Journal of Regional Science 11:317-326.

Nystuen, J.D. 1967. A theory and simulation of urban travel. In Garrison and Marble, eds. Quantitative Geography, Part I: Economic and Cultural Topics. Department of Geography, Northwestern University, Evanston, Illinois, pp. 54-83.

Northeast Corridor Transportation Project. 1970. Impact Studies: INTRA II Report NEC

Oates, W., W. Howrey and W.J. Baumol. 1971. The analysis of public policy in dynamic urban models. Journal of Political Economy 79:142-153.

Olsson, G. and S. Gale. 1968. Spatial theory and human behaviour. Papers, Regional Science Association 21: 229-242.

Operations Research. 1972. Urban problems: special issue. Journal of the Operations Research Society of America 2:3:

Paelinck, J. 1970. Dynamic urban growth models. Papers, Regional Science Association 24:25-38.

Paelinck, J. 1972. Formal allocation models for urban activities. Review of Regional Studies 2:

Pilgrim, B. 1969. Choice of house in a new town. Regional Studies 3:3: 325-330.

Prais, S.J. 1955. Measuring social mobility. Journal of the Royal Statistical Society. Series A. 118:56-66.

Putman, S.H. 1967. Intraurban industrial model design and implementation. Papers and Proceedings of the Regional Science Association 19:199-214.

Pyke, R. 1961. Markov renewal processes: definitions and preliminary properties. Annals, Mathematical Statistics 32:231-259.

Richardson, H.W. 1973. A markov chain model of interregional savings and capital growth. Journal of Regional Science 13:1:17-28.

Regional Planning Association of New York. 1962. Spread city. Bulletin 100. New York.

Robinson, I.M., N.B. Wolfe, and R.L. Barringer. 1965. A simulation model for renewal programming. Journal, American Institute of Planners 32:126-133.

Rogers, A. and R. Miller. 1967. Estimating a matrix population growth operator from distributional time series. Annals, Association of American Geographers 57:751-756.

Rogers, A. 1968. Matrix Analysis of Interregional Population Growth and Distribution. Berkeley: University of California Press.

Rose, H.M. 1970. The development of an urban subsystem: the case of the negro ghetto. Annals, Association of American Geographers 60:1-17.

Rose, H.M. 1970. The structure of retail trade in a racially changing trade area. Geographical Analysis 2:135-148.

Ross, P.S. and Partners. 1972. Promus: A Simulation System for Urban Analysis and Management. Toronto.

San Francisco Department of City Planning. 1968. Status of the San Francisco Simulation Model. San Francisco.

Schlager, K. J. 1964. Simulation models in urban and regional planning. S.C. Wisconsin Regional Planning Commission Technical Record 11, No. 1. Wankesha, Wisconsin.

Schlager, K.J. 1965a. A land use plan design model. Journal, American Institute of Planners 32:2:103-111.

Schlager, K.J. 1965b. Recursive programming model of the residential land development process. Highway Research Record.

Scott, A.J. 1969. Dynamic location - allocation systems: some basic planning strategies. Discussion Paper No. 3. Department of Geography, University of Toronto.

Scott, A.J. 1971. Optimal decision processes for a class of dynamic locational problems. Papers, Regional Science Association 26:25-36.

Sears, D.W. 1971. Computer simulation of a regional housing system. Papers on the Application of Computers to the Problems of Urban Society, 6th Annual Symposium. New York: Association of Computing Machinery.

Silvers, A.L. and A.K. Sloan. 1965. A model framework for comprehensive planning in New York City. Journal, American Institute of Planners 31:246-251.

Smith, P.E. 1961. Markov chains, exchange matrices, and regional development. Journal of Regional Science 3:27-36.

Smith, N.J. and A.P. Sage. 1973. Hierarchical system identification of models for urban dynamics. Socio-Economic Planning Sciences 7:5:545-569.

Spilerman, S. 1972a. Extensions of the mover-stayer model. American Journal of Sociology 78:599-627.

Spilerman, S. 1972b. The analysis of mobility processes by the introduction of independent variables into a Markov chain. American Sociological Review 37:277-294.

Steger, W.A. 1964. Simulation model. Community Renewal Planning, Progress Report 5. Pittsburgh City Planning Department.

Steger, W.A. 1965. The Pittsburgh urban renewal simulation model. Journal, American Institute of Planners 32:2:144-150.

Steinitz, C. and P. Rogers. 1970. A Systems Analysis Model of Urbanization and Change. Cambridge, Mass.: M.I.T. Press.

Stone, L.O. 1968. Stable migration rates from a multi-regional growth matrix operator. Demography 5:439-442.

Stone, R. 1972. A Markovian education model and other examples linking social behaviour to the economy. Journal of the Royal Statistical Society, A. 135:511-543.

Stonebreaker, M. 1972. A simplification of Forrester's model of an urban area. IEEE Transactions on Systems, Man, and Cybernetics SCM-2:4:468-472.

Stoyle, J. 1963. Land utilization as a stochastic process. Canadian Journal of Agricultural Economics 11:2:52-64.

Swanson, C.V. and R.J. Waldmann. 1970. A simulation model of economic growth dynamics. Journal, American Institute of Planners 36: 316-322.

Sykes, Z.M. 1969. Some stochastic versions of the matrix model for population dynamics. Journal of the American Statistical Association 64:111-130.

Taaffe, E.J., B. Garner and M. Yeates. 1963. The Peripheral Journey to Work: A Geographical Consideration. Evanston: Northwestern University Press.

Tapiero, C. 1971. Transportation-location allocation problems over time. Journal of Regional Science 11:377-384.

Tarver, J.D. and W.R. Gurley. 1965. A stochastic analysis of geographic mobility and population projections of the census divisions in the United States. Demography 2:134-139.

Theil, H. 1972. Statistical Decomposition Analysis with Applications in the Social and Administrative Sciences. New York: American Elsevier.

Tintner, G. and J.K. Sengupta. 1972. Stochastic Economics: Stochastic Processes, Control, and Programming. New York: Academic Press.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234-240.

Tobler, W.R. 1970. Review of urban dynamics. Geographical Analysis 2:198-199.

Toregas, C. and C. Revelle. 1972. Optimal location under time or distance constraints. Papers of the Regional Science Association 28:133-144.

Trinkl, F.H. 1973. Hierarchical resource allocation decisions. Policy Sciences 4:211-221.

Usher, M.B. 1972. Developments in the Leslie Matrix Model. In J.N.R. Jeffer , ed. Mathematical Models in Ecology. 12th Symposium of the British Ecological Society. Blackwells. pp. 29-60.

Vancouver Regional Simulation Study. 1972. First and Second Year Reports - 1971, 1972. Vancouver: Resource Science Centre.

Veinott, A.F. 1966. The status of mathematical inventory theory. Management Science 12:11:745-777.

Wagner, H.M. 1969. Principles of Operations Research. Englewood Cliffs: Prentice-Hall.

Walsh, B.F. and S. Grava. 1969. A dynamic land-use allocation model. Paper on the Application of Computers to Problems of Urban Society, 4th Annual Symposium. New York: Association of Computing Machinery.

Wärneryd, Olof. 1968. Interdependence in urban systems. Meddelanden Fran Göteborgs Universitets Geografiska Institutioner, Ser.B no.1.

Weathersby, G.B. 1970a. The allocation of public housing resources: an application of decision and control theory. Department of Economics, University of California at Berkeley. Mimeographed.

Weathersby, G.B. 1970b. Formal models of filtering in the housing stock. Department of Economics, University of California at Berkeley. Mimeographed.

Weeda, P.J. 1970. A production problem with a non-denumerable state space. In DeLeve, G., et al. Generalized Markovian Decision Processes. Amsterdam: Mathematical Centre, pp. 78-107.

Wheaton, W.C. and B. Harris. 1971. Linear Programming and Residential Location: The Herbert Stevens Model Revisited. Planning Sciences, University of Pennsylvania.

White, H.C. 1971. Multipliers, vacancy chains, and filtering in housing. Journal, American Institute of Planners 37:88-94.

Williamson, J.G. and J.A. Swanson. 1966. The growth of cities in the American Northeast, 1820-1870. Explorations in Entrepreneurial History 4:1

Wingo, L. 1961. Transportation and Urban Land. Washington: Resources for the Future.

Wolfe, H.B. 1967. Models for conditioning aging of residential structures. Journal, American Institute of Planners 33:192-196.

Wolfe, H.B. 1967. Model of San Francisco housing market. Socio-Economic Planning Sciences 1:71-95.

Wolfe, R. and G.B. Dantzig. Linear programming in a Markov chain. Operations Research 10:702

APPENDIX B

Three Annotated Bibliographies

For purposes of the initial reviews of the rather large literature on urban change the three student assistants were assigned areas of research which were somewhat arbitrary but seemed to have some validity and roughly corresponded to their interests at that time. The result of their labours appears in this appendix and in Chapter II.

Part I is an annotated bibliography of urban models and urban-related models which use the Markovian framework of analysis. Its companion review paper appears as Chapter II of the study. Part II is an annotated bibliography of urban simulation models together with an accompanying review paper. Finally, Part III has as its terms of reference economic models of urban growth.

It is hoped that these reviews will be a useful summary and interpretative guide to policy makers, planners, and researchers.

APPENDIX B

Part I: Markov Models

Russell Lee

Adelman, I.G. 1958. A stochastic analysis of the size distribution of firms. Journal of the American Statistical Association 53: 893-904.

System states in this study were firm size classes, as measured by value of output. A transition matrix was computed from data on 1929-39 transitions; and there was a graphical and tabular comparison of the corresponding equilibrium vector and the 1956 distribution. Three problems that can be noted are:

i) the rather unsatisfactory estimation of transition probabilities - common to urban and regional analysis due to data inadequacies;

ii) the use of the equilibrium vector for comparison with the 1956 distribution; there is no guarantee that convergence to equilibrium occurs so rapidly; and

iii) the change in the definition of size classes over time to account for inflation and technology factors (industrial advances); this manipulation has some appeal, but is rather *ad hoc*.

Also an index for industrial mobility for time n is defined:

$$I_n = \frac{\sum_{j=o}^{m} \frac{t_j}{1 - t_j}}{\sum_{j=o}^{m} \frac{S_{j,n}}{1 - P_{jj}}}$$

where t_j is j^{th} element in equilibrium vector

$S_{j,n}$ the j^{th} element in distribution vector at time n

P_{jj} the jj^{th} element in transition matrix.

* * *

Blumen, I., M. Kogan and P.J. McCarthy. 1955. The Industrial Mobility of Labour as a Probability Process. Ithaca: Cornell University Press.

A study of the flow of manpower between occupational categories. Its main importance lies in its treatment of one aspect of population heterogeneity via the "mover-stayer" model. "Stayers" were regarded as not changing their jobs at all. "Movers" were all considered to behave according to the same decision process, represented by a Poisson distribution. This may be seen to be a special case of the semi-Markov model. This revised model produced a good fit when compared to the actual transition matrix, whereas the usual Markov model underestimated the main diagonal elements.

* * *

Bourne, L.S. 1969. A Spatial land use conversion model of urban growth. Journal of Regional Science 9:261-272.

The approach here employed a system of multiple regression equations to compute what can be considered as the initial state vector of land use categories. The transition matrix (an average over two time periods) was then used for forecasting the future distribution. It was suggested that successive iterations of this method may be used for forecasting. In the Markov framework, however, successive regression steps would be unnecessary, since only the initial vector is required. On the other hand, if the regression step is to allow for successive feedback effects, the forecasting power is limited to only one time period - unless estimates of the future values of the independent variables are estimated as well.

* * *

Bourne, L.S. and C.A. Maher. 1970. Comparative extrapolations of city-size distributions: the Ontario-Quebec urban system. Research Report No. 27. Toronto: Centre for Urban and Community Studies, University of Toronto.

Transition matrices for five and ten year intervals were constructed to forecast city size class distributions. Arbitrary entry figures were incremented to allow for new cities. Also, transition matrices were tested for stationarity using the chi-square and mean discriminant information tests.

* * *

Brown, L.A. and F.E. Horton. 1970. Functional distance: an operational approach. Geographical Analysis 2: 76-82.

Mean first passage time (MFPT) is used as an index of functional distance. Three properties of spatial interaction links are noted: indirect, as well as direct, effects; assymmetry; and hierarchical characteristics. The index used, however, is merely a summary description. Three problems are noted: large variance in MFPT's; absorbing chains are inadmissable; only a single flow characteristic (migration, here) is described.

* * *

Brown, L.A. 1970. On the use of Markov chains in movement research. Economic Geography 46: 393-403.

Reviews Markov chain applications in spatial analysis. Discusses uses of Markov chains as a descriptive technique, methods of estimating the transition probabilities, and problems of the model.

A Markov decision-making approach, following Howard (1960), is discussed. The illustrative example of maximizing "organization" (or "negentropy") is difficult to rationalize however.

Suggested modifications of the Markov model included:

i) a semi-Markov framework (Ginsberg, 1971);

ii) use of time-dependent variables in estimating transition probabilities (Bourne, 1969);

iii) introducing additional matrix operators to represent "contagion" effects (Olsson and Gale, 1968).

* * *

Clark, W.A.V. 1965. Markov chain analysis in geography: an application to the movement of rental housing areas. Annals, Association of American Geographers 55: 351-359

Markov chain analysis was applied to changes in average rents of rental properties by census tract. However, the analysis was limited to visual inspection of the transition probability matrices (for different American cities). Also, it did not seem clear whether the transition matrices were calculated from data from 1940-1950, 1950-1960, or both. And there were no tests for stationarity. As well, the aggregation necessary to avoid absorbing states seemed particularly "ad hoc." On the other hand, this was the seminal paper on the application of Markov chains to urban geography (to my knowledge); and many of its suggestions have since been followed up.

* * *

Cohen, J.E. 1972. Markov population processes as models of primate social and population dynamics. Theoretical Population Biology 3: 119-134.

Deals with transition rates which depend nonlinearly on the sizes of the origins and destinations, hence improving on both the linear one-step Markov chain, and birth-death-migration models which treat the components independently. The most general of the three models presented by Cohen is the system defined by:

(1) $\alpha_i(n_i) = a_i + b_i n_i$ $\qquad a_i b_i, d_i > 0$ and $b_i < d_i$

(2) $\beta_i(n_i) = d_i n_i$

$$(3) \quad \gamma_{ij}(n_i, n_j) = g_{ij}\ \beta_i(n_i)\ \alpha_i(n_j)$$
$$= g_{ij}\ (d_i n_i)\ (a_j + b_j n_j)$$

where
n_i number in state i
$\alpha_i(n_i)$ arrival rate
$\beta_i(n_i)$ departure rate
$\gamma_{ij}(n_i, n_j)$ rate of migration from i to j
a_i, b_i, d_i parameters
g_{ij} a measure of distance from i to j

Found that the marginal probability density function (i.e. distribution of the number in each state) is negative binomial. The "aggregate distribution" of the number of states containing s individuals at equilibrium was found to be negative binomial as well.

Essentially, the model presents a recursive scheme that invokes non-stationary arrival, departure, and migration rates; and sets them to be a function of the sizes of the states (origins and destinations). On the stationary or equilibrium conditions were dealt with, however. In an empirical example, one of the models was compared to an observed distribution using the χ^2; and the parameters $a_i = a$, $b_i = b$, $d_i = d$ were estimated using regression analysis. The upshot is that this was limited to a static approach, lacking any temporal estimates of n_i, a_i, b_i, and d_i.

* * *

deCani, J.S. 1961. On the construction of stochastic models of population growth and migration. Journal of Regional Science 3:1-13.

Outlines three stochastic models in which the transition probabilities are a function of the number (population) in each state. Several difficulties may be noted:

(1) due to analytical complexities, only the two-state case was considered;

(2) transition probabilities were only defined for a limited number of possible transitions (-1, 0, +1); this may be rationalized, however, by allowing the time interval h to approach 0;

(3) the specification of the transition probabilities in the pure migration and birth-death-migration models implied that the individuals in the system tend to migrate to the state with the smaller population; this is a rather unattractive property in view of traditional gravity and potential concepts;

(4) the predator-prey model appears suitable for urban filtering and invasion-succession type processes, but unsuitable for more general processes.

Denton, F.T. 1973. A simulation model of month-to-month labour force movement in Canada. *International Economic Review* 14: 293-311.

State variables in the transition matrices (one for males, one for females) denoted employment, not in labour force, and various states of unemployment. The parameters were estimated from econometric equations of the form, with corrections for autocorrelation where necessary.

$$P_{ij}, t = a_{ijo} + \sum_{k=1}^{11} a_{ijk} D_k + a_{ij12} \bar{U}_t + a_{ij13} \Delta U_t,$$

where D_k is a dummy variable representing the months of the year except December (December not included so as to avoid problem of matrix singularity).

$\bar{U}_t$ a moving average, $\frac{1}{2}(U_t + U_{t+1}$

ΔU_t first-order difference, $U_{t+1} - U_t$

U_{t+1}, U_t unemployment rate at times (t+1) and t (these were not used in the regression estimation due to the high correlation between them.

The model has been extended by disaggregating by age, and by region - although these results were not reported. The main shortcoming of the model appears to be the fact that its use is restricted to simulation rather than prediction, Prespecified values of U_{t+1} are required since these are not estimated endogenous to the model.

* * *

Drewett, J.R. 1969. A stochastic model of the land conversion process: an interim report. Regional Studies 3: 269-280.

This study presents a simulation of the suburbanization process, the state variable being expressed in terms of percentage of land devoted to urban activity in a 1 km. square. The simulation was achieved by two drawings of random numbers. One, to define the waiting time in a state i, given in continuous time parameter space by

$$te = \frac{\log_e (1/U)}{M_i}$$

where U is a random number, $0 < U < 1$;

and $$M_i = \frac{-\log_e P_{ii}}{\Delta t}$$, and Δt is set to be unity.

Second, to determine the actual transition from an appropriately defined transition matrix, P, which is estimated empirically.

Comments

As with simulation models in general, the results should be compared to the actual results from another time period - otherwise the results may simply be duplicating those observations used in the estimation of the model. Additionally, the use of the binary variable urban (non-urban) land use prevents differentiating among varied land use types. Hence, this is obviously an incomplete model of aggregate urban structural evolution.

* * *

Fano, P.L. 1969. Organization, city size distributions and central places. Papers, Regional Science Association 22: 29-38.

Suggests the Markov model as an alternative to entropy maximizing approaches to city size distribution. However, as noted later, entropy can be used as a measure of disorder in a process represented by a Markov chain, so that the Markov model is not at all an alternative. As well, Fano notes that heterogeneity will alter the equi-probable solution of the entropy maximization problem; and a rank size rule example is given.

* * *

Fisher, J.C. and B.R. Lawson. 1972. Spatial planning in Yugoslavia - an application of Markov chain analysis to changing settlement patterns. The Review of Regional Studies 2: 195-214.

Transition matrices estimated from data from two time periods were suggested to be useful for description of the past as well as for prediction. The states were defined by population size intervals of settlements and also of grid cells; the size of the intervals were not equal and appeared arbitrarily designed. A scheme with additional states to allow for different potentials of entry (i.e. development of a settlement) and exit was suggested, but not operationalized.

Fuguitt, G.V. 1965. The growth and decline of small towns as a probability process. American Sociological Review 30: 403-411.

Applies the Markov chain model, with states being (unequal) population intervals for small towns (largest size category was "5000 and over"). Data at the beginning and at the end of each decade (for which data existed at the beginning) were used to compute the transition probabilities, for eight different decades from 1880 - 1960. Not unexpectedly, the transition matrices exhibited a trend toward larger size categories, with greatest movements in the years prior to 1900 (when the average size of towns was smaller than at present), and least from 1920-30 (during the depression). Also, it was concluded that the transitions were not stationary, although this was not tested formally.

The distributions of sizes for every ten years were compared when:

(a) incorporations and discorporations of towns were added or subtracted as they occurred, and the actual transition matrices at each interval were used; and

(b) identical treatment of incorporations, but with two different extreme transition matrices.

Of course the results in (a) was between the two extremes in (b), so that the point of the exercise seemed unclear.

Another comparison was made between the results in (a) and:

(c) the results of imputing the average incorporations from the earlier decades into estimates at later decades using the appropriate transition matrices of the later decades. Case (c) resulted in a greater proportion of smaller towns, and smaller proportion of larger towns than case (a). The result is obvious from the fact that there were more incorporations in the earlier decades.

In general, this was a very limited analysis based almost entirely on visual inspection.

* * *

Gale, S. 1972. Some formal properties of Hagerstrand's model of spatial interaction. Journal of Regional Science 12: 199-217.

Notes that Hagerstrand's Mean Information Field (MIF) can, under certain conditions, be regarded as isomorphic to a transition probability matrix, and hence as a special case of a Markov model. In this context, several points are discussed:

(1) the isomorphism between transition matrices and contingency tables allows application of techniques developed for the latter;

(2) the estimation of the transition probabilities, including the Bayesian approach;

(3) greater specification of the substantive interpretation of the states and transition probabilities; this has been studied in more detail by Ginsberg (1971, 1972a, 1972b);

(4) methods for incorporating a birth-death component into what is basically a migration model:
  i) simulation in which behavioural interpretation is limited;
  ii) adding a row and column to the transition matrix, to represent birth and death probabilities; this creates interpretation problems in that one state simultaneously corresponds to both births and deahts;
  iii) adding birth and death matrices, which must be determined exogenously;
  iv) allowing the row sums of the transition matrix to be other than one, which is somewhat analogous to (ii), but without some of the interpretation problems.

* * *

Gilbert, G. 1972. Two Markov models of neighbourhood housing turnover. Environment and Planning 4: 133-146.

A non-homogeneous Markov model and a Markov renewal process model are presented as less restrictive alternatives for the Markov chain model, in the context of housing turnover.

The non-homogeneous Markov model includes the effects of neighbourhood composition on turnover probabilities, in that the latter are a function of the time t. The main result for the two-state case is,

$$E\ [N(t)] = j\ e^{-2p(t)}$$

where $E\ [N(t)]$ expected no. housing units in state 1 at time $t \geq 0$

j no. units initially in state 1

$p(t) = \int_0^t [\mu(t) - \lambda(t)]\, d\tau$

$\mu(t), \lambda(t)$ entry and exit probability functions for state 1.

However, this solution requires that inferences be made about state 0 from information of state 1, so that a two-state system is unavoidable.

In the Markov renewal process, the main relation is:

$$A_{ij}(t) = \tilde{A}_{ij}\ F_{ij}(t)$$

where $A_{ij}(t)$ semi-Markov matrix, giving the probability of i - j move at $\leq$ time t, given move to i at t=0.

$\tilde{A}_{ij}$ probability of an i - j move at all

$F_{ij}(t)$ probability of an i - j move at or before t, given i - j move; i.e. the distribution function of occupancy times in i, given next state to be j.

From this, Gilbert develops expressions for the expected number of units in each state, turnover rate, equilibrium vector, first passage times, number of entries into each state, and expected sojourn times in each state; as well, some numerical two-state examples are presented.

In the non-homogeneous model the exponential parameter of the occupancy time distribution is allowed to vary (it is constant in the homogeneous Markov model) ; in the renewal model, any form of $F_{ij}(t)$ is allowed. Hence, the non-homogeneous model is a special case of the renewal model. It is surprising, then, to note that the results derived for the former are explicitly restricted to the two-state case; whereas this is not true in the latter. In addition, calibration of the renewal model requires estimates of

$\tilde{A}_{ij}$ and $F_{ij}(t)$, which may be obtained directly by empirical measurement of longitudinal data, records of real estate transactions, etc.; whereas in the other model, the form of the parameters $\chi_n(t)$ and $\mu_n(t)$ must be assumed.

Although an improvement over the Markov chain model it should also be realized that these models account for only a portion of the non-homogeneous effects. Also, the Markov renewal process strictly holds only for transitions which occur independently. At the individual housing unit level, this is a rather severe assumption when one considers housing turnovers due to large development and redevelopment projects.

* * *

Ginsberg, R.B. 1971. Semi-Markov processes and mobility. Journal of Mathematical Sociology 1: 233-262.

A Semi-Markov model of migration is presented; and some of its properties derived. The model is a generalization of the Markov process in that, in addition, the probability of leaving a state can depend on the length of time that state has been occupied (duration-of-stay effect) and on the next state entered (pull, in addition to push effect). The 'cumulative inertia' axiom of McGinnis (1968) is shown to be accommodated within the Semi-Markov framework; and the 'mover-stayer' model of Blumen, Kogan, and McCarthy (1955), closely related. An outline of the requirements for operationalizing the model is given.

Outline of Model

Relevant notation is appended to this review. All equation numbers correspond to Ginsberg's. Assume that the probability that the n^{th} state entered is k and the n^{th} waiting time is $\leq X$, depends only on the $(n-1)^{st}$ state and the $Q_{ij}(t)$ distributions (denoting the probability of an i - j move at or before t, given move into i at time 0):

$$P(J_n=k,\ X_n \leq X\ |J_0,J_1,X_1,\ \ldots,\ J_{n-1},X_{n-1}) = QJ_{n-1}k^{(X)}. \qquad (21)$$

Let the initial distribution vector be A

$$A = (a_1, a_2, \ldots, a_n), \quad a_k = P(J_0 = k). \tag{22}$$

Define $N(t)$ = Supremum $\{n$, where $n \geq 0 \mid S_n \leq t.\}$

This represents the number of moves in the time interval $(0,t)$. Also define $N_j(t)$ to be the number of times $J_k = j$ for $0 < k < N(t) + 1$ (this excludes Jo). Clearly, $\sum_j N_j(t) = N(t)$.

What we have then is really a frequency distribution of destinations. And the counting process $\bar{N}(t) = (N_1(t), N_2(t), \ldots, N_n(t))$, which is specified by (n, A, Q), is called a Markov Renewal Process.

Let $Z(t) = J_{N(t)}$, the state occupied at time t. The sequence of states $\{Z(t): t \geq 0\}$ (correcting for the typographical error in Ginsberg) is called a Semi-Markov Process. And the process is characterized by $Q_{ij}(t) = P_{ij}F_{ij}(t)$, the probability of an i - j transition at time $\leq$ t, given a move to i at time 0.

The following lemmas follow directly from equation (21):

$$P(J_n = k,\ S_n \leq y)\ J_0, J_1, \ldots, J_{n-1}, S_{n-1}) = QJ_{n-1}k(y - S_{n-1}) \tag{23a}$$

(correcting for the typographical error in Ginsberg);

$$P(J_n = k \quad J_0, J_1, \ldots, J_{n-1}) = PJ_{n-1}k, \tag{23b}$$

where $P = \{p_{ij}\}$ is the transition Matrix of the corresponding Markov chain (CMC) of the Semi-Markov process (S-Mp), and is assumed stationary. From equation (23a), it is easy to see that the time between moves, S_n, depends on both J_{n-1} and $J_n = k$, the initial and final states. In the Markov model, on the other hand, S_n is independent of the destination.

Basic Properties of the Model

The main results of the derivations are given here. Details of the derivations in Ginsberg are generally easy to follow. The basic pattern in the proofs is to partition the process at the times when new states are entered, obtaining a sum of independent random variables. Distributions of these sums are then given by convolution integrals. Now, taking the Laplace-Stieltjes (L-S) transform of a convolution is equal to the product of the L-S transforms of the terms. So, simple products are substituted for complicated integrals; and appropriate manipulations give the desired results.

(1) Transition Probabilities:

i) $P(t) - (I-Q(t))^{-1} * (I-H(t)),$ (39)

where convolution replaces ordinary matrix multiplication; or in terms of the first passage times $G_{ij}(t)$,

ii) $P(t) = (I-H(t)) * (I-G(t))^{-1}.$ (42)

(2) Renewal Function:

The renewal function $W(t) = \{w_{ij}(t)\}$, the mean number of moves into j in (0,t) given Z(0) = i, is

$$W_{ij}(t) = G_{ij}(t) + \int_0^t G_{ij}(t-u)\, d\, W_{jj}(u). \tag{47}$$

Further, $W(t) = Q\,(t) \quad * \quad (I-Q(t))^{-1},$ (49)

$Q(t) = W\,(t) \quad * \quad (I+W(t))^{-1},$

so that knowledge of W is equivalent to knowledge of Q.

(3) Asymptotic Behaviour:

The mean recurrence time, $L_{ij} = E\,(G_{ij}(t)),$ is

$$L_{ij} = \sum_{k \neq j} p_{ik} L_{kj} + M_i \tag{53}$$

and,

$$L_{ii} = \frac{1}{\Pi_i} \sum_{k=1}^{n} \Pi_k M_k , \tag{54}$$

where, correcting for the typographical error in Ginsberg,

$\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_n)$ is the limiting vector of the P matrix

of the CMC, assumed regular. Further, if P is absorbing and $N = \{ n_{ij} \}$ is its fundamental matrix, the mean time until absorption, given i transient, is

$$n_i' = \sum_j n_{ij}\, m_j \tag{55}$$

And, if P is regular,

$$\lim_{t \to \infty} P_{ij}(t) = \frac{m_j}{L_{jj}} \tag{56}$$

$$\lim_{t \to \infty} W_{ij}(t) = \frac{1}{L_{jj}} \tag{57}$$

From these results (equations 53 - 57), it is clear that many of the basic

quantities of semi-Markov processes are the same as those in their CMC, except that they are weighted by the mean transition times M_{ij} and M_i.

Operationalizing the Model

To compute the principle attribute of the process, $Q_{ij}(t) = P_{ij}\, F_{ij}(t)$, one must estimate the parameters:

a) $P = \{P_{ij}\}$ of the CMC; and the

b) $F_{ij}(t)$'s.

These values allow computation of other properties, as noted. Note that, for many cases, the specific form of the $F_{ij}(t)$'s is unnecessary, since only their means, M_{ij}'s, enter into the pertinent expressions.

In addition, to test the tacit assumptions of stationarity of P and independence of times between successive moves, the Anderson and Goodman (1957) and Cox and Lewis (1966) tests can be used, respectively.

Relationship to Other Models

Ginsberg notes that if F(t) is chosen to be Gamma with parameter $\alpha < 1$, the duration-specific "decreasing failure rate"property of the McGinnis (1968) cumulative inertia axiom, is satisfied.

Also, since the mover-stayer model partitions the population into stayers, with probability of moving 0; and movers, governed by a Markov chain - its relationship with the Semi-Markov model is seen.

Extensions of the Model: Non-homogeneity

(1) Age Effects:

Age, or individual's life cycle stages, influence the probability of moving. Hence, as the process develops, the transition mechanism is non-stationary. However, if the interactions between age and location are not too great, age can be accommodated within a homogeneous framework using a new time scale, τ, which Ginsberg calls operational time. In effect, the time scale is stretched (or contracted) so that the mean number of events in (0,t) is always τ:

$$W(t) = \tau$$

(2) The First Move:

It can be postulated that there may be a difference between the probability distribution of the first move out of the parental home, and that of subsequent moves. The resultant modified renewal process can be treated by partitioning the process into the first move, and the delayed recurrent events (subsequent moves). In this case, it appears that

$$W(t) = \hat{Q}(t) \quad * \ (I-Q(t))^{-1} \tag{69}$$

(a correction of Ginsberg's equation 69), where $\hat{Q}(t)$ pertains only to the distribution of the first move.

Comments

The generality of the semi-Markov model relative to the Markov, is readily apparent - the latter being a special case of the former. Yet, it is still a device for handling a system which we do not understand well enough to treat by means of differential or difference equations (Bellman, 1965). The model simply transforms the transition matrix, P of the CMC, by matrix operators, F(t). P must be estimated empirically, which is the same requirement as for the ordinary Markov chain situation. Now, however, F(t) must be estimated as well. It appears that in fact the general form of F(t) is to be assumed - for example, a Gamma distribution. Hence, a distinctly Bayesian flavour appears - specifically, the subjective selection of the form of F(t). The same observation is relevant to Ginsberg's discussion of operational time, in which the form of W(t) must be assumed (p.258). In this particular context, the form $W(t) = \frac{\alpha}{\beta}(1-e^{-\beta t})$ suggested (p.258), requires that the parameters α and β must be simultaneously evaluated from the data -- requiring heuristic calibration techniques.

Little attention is given to parameter estimation problems. For example, in a ten region system and if the $F_{ij}(t)$ are assumed to have two parameters, about 300 parameters must be estimated in toto (Ginsberg, 1971). Consequently, this implies the requirement of either: (a) a large sample; or (b) simplification of the $F_{ij}(t)$, for example letting $F_{ij}(t) = H_i(t)$. If a suitably large sample is not available, then the $F_{ij}(t) = H_i(t)$ suggestion requires that the probability of an i - j transition does not depend on j, the destination. Further, estimation of P using exogenous

information such as economic factors, is constrained by data limitations, as well as the usual problems related to regression analysis. Perhaps a better tack would be to avoid the $F_{ij}(t) = H_i(t)$ restriction; and substitute Bayesian estimation methods for strictly maximum-likelihood estimates - for example using a general Dirichlet distribution, if a multiple sample is available (Good, 1965).

Ginsberg's treatment of non-homogeneity by considering age effects is grounded on the assumption that interactions between age and location are insignificant. However, the strong counter-examples provided by Ginsberg himself (e.g. people are more likely to move to cities when they are young) (pp.258-259), tend to refute this assumption.

Certainly, many other sources of non-homogeneity remain to be treated -- economic factors; the effect of previous mobility history <u>before</u> J_{n-1}, on a transition to J_n (this would appear to render the model non-Markovian); social factors and communication of information among acquaintances. Some of these are considered in subsequent papers by Ginsberg.

<u>Notation</u>

$*$		convolution operator
Π	=	$(\Pi_1, \Pi_2, \ldots, \Pi_n)$, the limiting vector of P of the CMC
τ		operational time
A		$A = (a_1, a_2, \ldots, a_n)$; $a_k = P(J_o = k)$, initial distribution vector
$F_{ij}(t)$	=	$\frac{Q_{ij}(t)}{P_{ij}}$, probability of i - j move at or before t, given some i - j move.
$G_{ij}(t)$	=	$P(N_j(t) > 0 \mid (Z(0) = i)$, first passage time
$H_i(t)$		probability of leaving i to any j at or before t $= \sum_{j=1}^{n} Q_{ij}(t)$
J_k		k^{th} state occupied
L_{ij}		Mean recurrence time
M_i		$E(H_i(t))$
M_{ij}		$E(F_{ij}(t))$
n		number of states
n_i		mean time to absorption, given i transient

$N(t)$ number of moves in $(0,t)$

$N_j(t)$ number of times $J_k=j$, for $0 < k < N(t) + 1$

$N_{ij}(t)$ no. times enter j in $(0,t)$, given $Z(0) = i$, so $W_{ij}(t) = E(N_{ij}(t))$.

N $= \{n_{ij}\}$, the fundamental matrix of an absorbing P of the CMC

P_{ij} probability of $i - j$ at all, $= Q_{ij}(\infty)$

$P(t)$, $P_{ij}(t)$, $P = \{p_{ij}\}$, the transition matrix of the CMC, the probability in j at time t given a move into i at t=0

$Q_{ij}(t)$ probability of $i - j$ move at $\leq$time t, given move to i at t=0

S_k time at which k^{th} state entered

$W_{ij}(t)$ mean no moves into j in $(0,t)$, given $Z(0) = i$

X_k time from $k\text{-}1^{st}$ to k^{th} state

$Z(t) = J_{N(t)}$, state occupied at time t.

* * *

Ginsberg, R.B. 1972. Critique of probabilistic models: application of the Semi-Markov model to migration. Journal of Mathematical Sociology 2: 63-82.

The thesis of this paper is that unless parameters of probabilistic models are related to the causal structure and exogenous determinants of the (migration) process, such models will have little substantive or theoretical content.

Limitations of Regression

Various limitations of regression analysis are discussed:

(1) static nature;

(2) difficulties in the specification of the error term; forcing total predicted flows from state i to sum to the actual total, causes the flows to be no longer random independent variables;

(3) the flows do not take into account any competition;

(4) although implicitly Markovian, non-Markovian properties are incorporated through parameter variation;

(5) describes only one characteristic of the process, viz. transition or migration over a fixed lag - what about first passage times, times between moves, etc.;

(6) although dealing with a complex system, there is usually no use of a set of simultaneous equations.

Comments on Use of Semi-Markov Model

The Semi-Markov approach preferred by Ginsberg is basically an aggregate model. The critical assertion here then, is that individuals' and aggregate social processes are equivalent aspects of the same random process connected by a matrix renewal equation. Hence, although not denying individual differences, he claims that some of these are spurious (p.69-70). This raises the question of what in fact is spurious. And would it not seem more plausible to posit the existence of a small set of groups of relatively homogeneous individuals? This would suggest the necessity of a set of renewal equations, if the variance in population attributes was sufficiently great. This in turn would multiply the already heavy data demands of the Semi-Markov model. Disaggregation in terms of age, sex, income, socio-economic status, ethnic status, and inter- vs. intra-metropolitan moves, compound the problem further. The number of case histories required would be of the order of 100,000. Even so, many of the elements in the empirical transition matrix will likely be zero so that the sample is still effectively small. From this, it is clear that Bayesian probability estimation methods would be required.

Comments on Policy Formulation

Application of dynamic programming to Semi-Markovian decision processes can be used to solve the following problem (Howard, 1963, 1970):

Given payoffs associated with i - j transitions, and payoffs which are functions of length of time spent in i (given that j will be the next state entered); and given that transition probabilities, transition times and the payoff structure, can be influenced through policy implementation -- which policy alternative maximizes total expected payoff over a specified planning horizon?

The effects of policy alternatives involving welfare payments, wage subsidies, public housing, etc., are realized through the parameters in the Markov renewal programming formulation. Specification of these relationships is a major task. Another requirement involves defining what in fact the payoffs of transitions themselves mean.

* * *

Ginsberg, R.B. 1972. Incorporating causal structure and exogenous information with probabilistic choice models: with special reference to choice, gravity, migration, and Markov chains. Journal of Mathematical Sociology 2: 83-103.

The paper discusses regression analysis and doubly stochastic estimating procedures for determining the elements of the transition matrix of the corresponding Markov chain $P = \{p_{ij}\}$, of the Semi-Markov model, developed in a previous paper (Ginsberg, 1971). Since these approaches are lacking in theoretical content, Luce's (1959, 1963) choice theory is introduced to model the structure of P. Initial P_{ij}'s estimated from data using maximum likelihood, are imputed into a weighted, generalized econometric regression scheme. The resultant regression coefficients (β's) may then be generally used to

estimate P_{ij}'s. If, however, the unexplained variance of the P_{ij}'s on the β's is greater than the variance in the data, a doubly stochastic model is required. An example of the latter occurs when the P_{ij}'s are derived by assuming them to be generated by a matrix beta distribution.

Ginsberg, however, considers these methods rather "ad hoc," and develops models of the structure of P based on Luce's choice theory. Hence, relative magnitudes of P_{ij} reflect determinants of individuals' choice process. The original theory distinguishes between determinants of intrinsic properties of response alternatives, and the relation of the alternatives with the individual's present state. A gravity model analog is constructed by relating the former concept with the intrinsic attractiveness of a destination, and the latter with distance between destinations and origin.

Specifically, Ginsberg, after Luce notes that <u>if</u> the probability of choosing alternative x from T is:

Axiom 1: $P_T(x) = P_T(R)\, P_R(x)$ for all $R \subset T$, $x \in R$ and $P_T(x) \neq 0,1$ (Ginsberg's assertion can be extended, for all $R \subseteq T$),
then there exist numbers b_x, unique up to a multiplicative constant, such that

$$P_R(x) = \frac{b_x}{\sum_{y \in R} b_y} \qquad \text{for all R.}$$

The term b_x, then, can be considered as a measure of the attractiveness of destination x; and is a function of socio-economic variables. To incorporate the distance element, an origin-specific modification of the attractiveness, f_i, is introduced. If it is accepted that (i) $f_i(b) > 0$, (ii) $b_x > 0$, (iii) $f_i(b_x)$ independent of $b_k \neq b_x$, and (iv) $f_i(kb) = kf_i(b)$, so that scale of b does not matter -- then the only permissible form of f_i is a diagonal matrix, resulting in multiplication of the b's by positive constants. Letting

$$f_i = \begin{pmatrix} a_{i1} & & & & 0 \\ & a_{i2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & \cdot & a_{is} \end{pmatrix}$$

and a_{ij} can be made a function of distance, so that a gravity type expression results.

$$P_{ij} = \frac{a_{ij}\ b_j}{\sum_k a_{ik}\ b_k}$$

Hence, the parameters that must be estimated are the b's and the a_{ij}'s (or equivalently, the parameter(s) of the distance function). Assuming the form of the distance function and specifying the determinants of b, allows the loglikelihood expression derived from the data to be solved for the b's and the parameters of the distance functions, using Newton's approximating procedure.

Comments

With respect to the regression analysis, the implicit assumption is that the β's are invariant over time, so that they can be used to forecast future p_{ij}'s, given the proposed determinants of the β's. Without confirmation, this conjecture is unacceptable.

The Luce formulation comprises the gist of Ginsberg's arguments; and is a means of incorporating causal structure into the model. Although the relative attractiveness expression

$$P_R(x) = \frac{b_x}{\sum_{y \in R} b_y}$$

may be intuitively obvious, it is in fact specified for these unique b_x (up to multiplication by a constant), if and only if Axiom 1 is satisfied. Hence from this conditional probability expression, directly interpretable within a choice framework, one gets an expression of relative attractiveness.

However, although the b_x are purported to imply a ratio scale (p.88), this degree of precision is unjustified in that Luce's choice axiom implies that even if alternative x is only slightly preferred to y, x will almost invariably be chosen (p.89).

It can be noted, too, that the form of f_i being a diagonal matrix, was directly determined by the 'a priori' assumptions. One of these was that $f_i(b_x)$ was independent of $b_{y \neq x}$.

This statement is somewhat disturbing in that it implies that there are no interactions between various destinations, overlooking such notions as spatial competition and intervening opportunities. Hence, Ginsberg appears to be over-reaching the capabilities of his model when he states that these concepts are directly interpretable from the formulation (p.92).

Finally, it should be noted that the Luce formulation is compatible with gravity type expressions; but it does not derive these. The actual form of the distance decay function must be assumed.

* * *

Henry, N.W., R. McGinnis, H.W. Tegtmeyer. 1971. A finite model of mobility. Journal of Mathematical Sociology 1:107-118.

The model presented here modifies the Cornell retention model (McGinnis, 1968) by placing a finite upper bound on the duration-of-stay effect, so that the probability of leaving alocation becomes constant after a certain duration time, r=v, at that location. This results in a finite Markov chain, which is either absorbing (if people will remain at that residence indefinitely) or regular (if there exists some constant non-zero probability of moving).

Absorbing Case

Important properties include:

(1) $N = (I-Q)^{-1}$, which gives the mean time spent in non-absorbing state j, given non-absorbing state i. Q is identified in the canonical form of the absorbing finite chain transition matrix

$$\begin{array}{ll} & \begin{array}{cc} \text{absorbing states} & \text{non-absorbing states} \end{array} \\ \begin{array}{l} \text{absorbing} \\ \text{non-absorbing} \end{array} & \left(\begin{array}{c|c} I & 0 \\ \hline R & Q \end{array}\right) \end{array}$$

N consists of ${}_{rs}N$ submatrices, giving mean number of times in a duration S location, before absorption, given an initial duration r location.

(2) The equilibrium distribution depends on the initial state, A(0): A(limit) = A(0)NR.

(3) The expected number of moves during a lifetime:

$$ {}_{ri}Ne = \begin{cases} (I-\prod_{k=r}^{v-i} {}_kS)M \quad {}_{11}Ne & r \quad 1 \\ (I-M+DM)^{-1}e & r = 1 \end{cases} $$

Where ${}_kS$ is diagonal matrix portion of ${}_kP$

M is the off-diagonal portion of ${}_kP$, assumed stationary

e vector of 1's

$$D \text{ equals } \prod_{r=1}^{v-1} {}_rS$$

<u>Regular Case</u>

The equilibrium distribution is:

$$A\ (\text{limit}) = \sum_{r=1}^{v} {}_rX$$

$$= {}_1X \left[I + \sum_{r=2}^{v-1} \left(\prod_{k=1}^{r-1} {}_kS \right) + \left(\prod_{k=1}^{v-1} {}_kS \right) (I - {}_vS)^{-1} \right]$$

where $X = ({}_1X, {}_2X, \ldots, {}_vX)$ and $X\Theta = X$, Duration $v \quad v-1 \ldots 2 \quad 1$

where $\Theta =$

$$\begin{array}{c} v \\ v-1 \\ v-2 \\ \cdot \\ \cdot \\ 2 \\ 1 \end{array} \begin{pmatrix} {}_vS & 0 & \ldots & 0 & {}_vM \\ {}_{v-1}S & 0 & \ldots & 0 & {}_{v-1}M \\ & {}_{v-2}S & & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & {}_2S & 0 & M \\ 0 & \ldots & & {}_1S & {}_1M \end{pmatrix}$$

a regular transition matrix.

<u>Comments</u>

"Duration-specific" classes multiply the number of parameters (transition probabilities) required, (proportional to r=V) and hence the data requirements and the manipulation problems. Further, the duration of residence effect is not at all well-specified; and is actually a form of disaggregation. All that is stated is that the transition probabilities differ in some way so as to satisfy the law of cumulative inertia. It is readily seen that the $F_{ij}(t)$ operator in Ginsberg's semi-Markov model (1971) is a generalization of the effects of r.

Also, it can be noted that the model is formulated for the discrete time case only.

And further, the introduction of a bound on the duration-of-stay effect guarantees a finite state space, hence allowing solutions. However, such a scheme is rather "ad hoc," and difficult to rationalize in the context of the actual process.

* * *

Hemmens, G.C. 1966. The structure of urban activity linkages. Centre for Urban and Regional Studies, Institute for Research in Social Science, University of North Carolina, Chapel Hill.

Provides a relatively non-technical discussion of the structure of urban activity linkages as revealed by intra-urban travel behaviour. Transition matrices, and related properties, were calculated for three cities, with the states being activity type for one case, and land use type in the other. Further applications which were suggested would consider variations among different cities, variations due to household characteristics, results from more detailed data, and time and spatial distributions related to the activities. In addition, the use of a semi-Markov model, more general than the Markov, was suggested. The following limitations (applicable to the Markov case as well) were noted:

i) the lack of an explicit spatial dimension,

ii) lack of causal relationships explaining individual behaviour, and

iii) the stationarity of the transition matrix, so that only short-run projections are allowed.

Subsequent suggestions by Ginsberg (1971, 1972a, 1972b) and others have begun to provide some extensions implicitly recommended by the above problems.

* * *

Harris, C.C., Jr. 1968. A stochastic process model of residential development. Journal of Regional Science 8: 29-39.

An attempt to simulate suburban residential development using a semi-Markovian framework (see Drewett, 1969). Several problems can be noted:

(1) Sub-areas in the fringe area undergoing development may be either developed (D) or not developed (N), and the states are defined by all possible arrangements of D and N among all sub-areas (this is essentially sampling with replacement). With m sub-areas, the number of states would be 2^m. In the Sacramento, California example, there were 73 sub-areas (p.36), giving 2^{73} states. But Harris is inconsistent in that he defines only m = 73 states (p.34,38). In fact, the "transition probabilities" he defines in the example are simply measures of relative accessibility for each sub-area; it is unclear how these relate to his former definition. Hence, there are two problems in this context: the great number of states, and the inconsistent treatment of the transition probabilities.

(2) Some arbitrary threshold on population density must be used to identify "developed." Allowing n, instead of 2, stages of development would require n^m states.

(3) Although purported to be a semi-Markov model, the operationalization in the example is unclear. An exponential (negative exponential?) function was supposedly assumed for the distribution of sojourn (length-of-stay) times, but the function was not made explicit. Further, the illustrative table (p.34) did not reflect an exponential (nor negative exponential) function.

(4) The treatment of accessibility is somewhat unusual. The distances to employment sites, for a given sub-area, were considered to have a multiplicative effect. This is questionable on two points:

i) defining the effects to be additive seems more justified; and

ii) it is unlikely that the different employment sites carry equal weights; only the closest one(s) will be of significance, otherwise why have any accessibility criterion?

In addition, although Wingo's (1961) study was consulted in defining accessibility, Harris omits the travel time factor - a salient aspect in Wingo's approach.

(5) The argument that the time for development to occur is independent of total fringe size (p.38) is unconvincing. For a given projected population growth and given sub-area, it would seem that the probability of it being developed if located in a two sub-areas fringe would be greater than in a 100 sub-areas fringe, other things being equal.

* * *

Harvey, D.M. 1967. Models in the evolution of spatial patterns in human geography. In R.J. Chorley and P. Haggett, eds. Models in Geography. London: Methuen, pp. 549-608.

An overview of quantitative models in human geography. The section on Markov chains uses a hypothetical 3-state example of household movement; and discusses previous applications.

* * *

Jones, C. 1971. Markov chains and their uses in urban and regional research. Occasional Paper No. 10. University of Manchester, Centre for Urban and Regional Research.

Summarizes many of the applications of Markov chains to urban and regional analysis.

* * *

Kelly, A.C. and L.W. Weiss. 1969. Markov processes and economic analysis: the case of migration. Econometrica 37: 280-297.

A Markov model of migration is compared to an economic model based on wage differences. The former is found to be plagued by the problem of non-stationarity; and by evidence that population changes before equilibrium is attained, are underestimated. This latter point was noted long before by Blumen, et al., 1955, who observed that main diagonal elements in the transition matrix underestimated observed values.

The Kelley-Weiss model is given by

$$\Delta C = \alpha \frac{W_e - W_r}{W_r} \text{ , which gives}$$

$$\frac{\Delta C}{C} = \beta \frac{W_c}{W_r} - \beta$$

where C is the proportion of the total population in California, for the two-region system, California-Rest of U.S.A.;
W_c, W_r wage rates in California and Rest of U.S.A., respectively;

$\alpha = \beta C$;

ΔC change in proportion of total population in California.

From $\Delta C = K_r R_{t-1} - K_c C_{t-1}$,

one gets $$\frac{dK_c}{dC} = \frac{-Kr}{C^2} + \frac{W_c}{W_r} \frac{1}{\eta(1-C)} + \frac{1}{\varepsilon_c} \text{ ,}$$

where $W_c = h_c C^{-(1/\varepsilon)}$

$W_r = h_r R^{-(1/\eta)}$

ε, η elasticities of demand for labour in California and rest of U.S.A., respectively

h_c, h_r constants

After estimating the orders of magnitude of the parameters and variables, dK_c/dC was evaluated and was found to be consistently less than zero. It can be suggested that much of the thrust of Kelley and Weiss' criticisms of the Markov model can be blunted by adopting a semi-Markov model. However, the economic model has the advantage of explicitly considering causal factors. These in turn could be accommodated within a semi-Markov framework - for example, by setting the Markov parameters to be a function of wage differences.

* * *

Keyfitz, N. 1972. On Future Population. Journal of the American Statistical Association 67: 347-363.

An overview of approaches to population projection and forecasting. "Projections" are viewed as being determined from a set of assumptions on births, deaths, and migration; "forecasts" are (unconditional) predictions.

Reviews cohort-survival model, and modifications by: (1) further disaggregation of the population; (2) changing the rates of birth and death, by defining them as trends, determined exogenously; (3) adding a migration component.

Outlines the forecasting methods of leading and lagging series, autoregressive series, and regression.

No explicit discussion of Markov chains.

* * *

Land, K.C. 1969. Duration of residence and prospective migration: further evidence. Demography 6: 133-140.

Following Morrison (1967), age-specific duration-of-stay effects were estimated for Monterrey, Mexico and Amsterdam, The Netherlands, using regression analysis. The conclusion was that the relationships in the two cases were comparable (i.e., "negative curvilinear"). Three problems may be noted however:

(1) the hypothesis holds only for a certain range of duration values (the parameters have unequal signs);

(2) as noted by Land, when comparing the two cases, the regression coefficient magnitudes differed significantly; in fact, the signs were often reversed.

(3) no R^2 values or t-statistics for the regression coefficients were given.

* * *

Lindsay, I. and B.M. Barr. 1972. Two stochastic approaches to migration: comparison of Monte Carlo simulation and Markov chain models. Geografiska Annaler 54B: 56-67.

An empirical comparison of Monte Carlo simulation and Markov chain models of interregional migration. Simulation was generated using the expression

$$E_{ji} = \frac{P_i H_i}{\sum_i P_i H_i} \quad B$$

where E_{ji} is probability of interaction from j to i

H_i population of destination j

$$B = \begin{cases} 0 & \text{barrier present} \\ 1 & \text{no barrier to interaction} \end{cases}$$

P_i smallest probability of contact between j and i, derived using a Pareto expression

For the Markov model, transition probabilities were estimated using a gravity expression, the parameters being calibrated from a questionnaire survey sample. And the equilibrium matrix was used to derive the expected values. For both models, the expected values derived were compared to the actual, calculated by a residual-type method using census data. Results of the Kolmogorov-Smirnov tests indicated that both models were reasonable; however, the Markov was favoured since it was more realistic and yet simpler.

In the Markov analysis, to allow for nonstationarity, a different transition matrix was computed for each time period. The transition matrices are hence derived a posteriori, so that the model in effect loses its forecasting power. Also, there is no reason to believe that the system is in equilibrium at each time period; in deriving the steady state distribution it is that which occurs in the long run, assuming stationarity. Hence, the computation of the equilibrium distribution for each time period is questionable.

* * *

Marble, D.F. 1964. Two computer programs for the analysis of simple Markov chains. Research Report No. 6. Department of Geography, Northwestern University, Evanston, Illinois.

Lists two computer programs for analyzing simple Markov chains. An application was given for intra-urban trip purpose linkages. Computations were made for the transition matrix, calculated from aggregated data on person movements; the equilibrium vector; and the mean and variance of the number of stops by purpose, depending on the first stop purpose. The equilibrium vector was compared to the actual distribution, for the time from which the equilibrium vector was calculated. This implicitly tested the existence of some sort of an equilibrium at that time, which in this context seems to imply some notion of stereotyped travel behaviour. It is worthwhile to note as well that it remains to be tested what time distribution existed between travel stops. In this context, the fact that the phenomenon is cyclical (the individual always returns home) and time-of-day dependent, poses additional questions.

* * *

Matras, J. 1967. Social mobility and social structure: some insights from the linear model. American Sociological Review 32: 608-614.

Suggests a model of the form:

$$a_o M = a_1$$

where a_i vector of socio-demographic structure at time i

M transition matrix, having fertility and social mobility components, disaggregated with respect to age groups and social classes.

* * *

McFarland, D.D. 1970. Intragenerational social mobility as a Markov process: including a time-stationary Markovian model that explains observed declines in mobility rates over time. American Sociological Review 35: 463-476.

Suggests a completely disaggregated Markov model of social mobility which is:

i) stationary

ii) has the Markovian property

iii) characterized by a heterogenous population.

In essence, each individual is characterized by its own transition matrix. The one-step population matrix would then be:

$$Q_1 = N_o^{-1} \sum_m N_o(m)\ P(m)$$

where $N_o(m)$ is a diagonal matrix with an entry of unity in the diagonal position corresponding to person m's initial state and o's elsewhere

$N_o \quad = \quad \sum_m N_o(m)$

P(m) person m's transition matrix

Q_1 one-step transition matrix for the population.

Comparable expressions for the k-step and equilibrium matrices were given and noted to be non-Markovian (this is due to the presence of the $N_o(m)$ matrices).

As claimed, this is a generalization of the mover-stayer model, but the latter may at least be operationalized. Certainly, as McFarland argues, the Cornell Model suggests only duration effects as an extension

of the Markov model. But the McFarland model, although superficially appealing in its conceptualization, runs into severe problems when one considers the astronomical parameter estimation demands. Aside from this, however, it is perplexing how individual transition probabilities could be measured. For example, does one ask an individual that if he was living in i what would be his probability of moving to j? This would be carrying heuristics to the extreme.

* * *

McGinnis, R. 1968. A stochastic model of social mobility. American Sociological Review 33: 712-722.

Exposition of the "Cornell Model," in which the Markov chain model is extended via the "Axiom of Cumulative Inertia." This latter postulate states that "the probability of remaining in any state of nature increases as a strict monotone function of duration of prior residence in that state." (p.716) This notion is operationalized by including an additional dimension (predictor variable) in the transition matrix. The model, however, is reduced to a Markovian formulation, so that a strict Markov chain characterizes the transitions within duration-specific classes.

This study presents a first step in relaxing the Markov chain model, and is an extension of the mover-stayer model (Blumen, et al., 1955); a further generalization has been outlined by Ginsberg (1971). The basis of the Cornell Model is that it effectively allows for adjustment of the main diagonal elements of the transition matrix. It should be kept in mind, however, that Markovian properties not affected by this additional axiom are retained; as an example, population heterogeneity remains unaccounted for. As well, it is important to note that the axiom as stated is an overgeneralization, and in fact misleading. The nature of the duration effect will be a function of the phenomenon under consideration.

* * *

Morrison, P.A. 1967. Duration of residence and prospective migration: a stochastic model. Demography 4: 553-561.

Empirical verification (using Amsterdam, The Netherlands data) of the Cornell mobility model (McGinnis, 1968), which stressed duration-of-stay effects on the propensity to migrate. The relationship was found to be negative (propensity decreases with increased duration) and curvilinear. In addition, analyses on age-specific matrices revealed that the propensity declines increased with age up to the "over 65 yr." group, at which the trend was not as strong.

An expression of the form

$$Y = a + b_1X + b_2X^2$$

where Y estimate of probability of migrating (at all)

X logarithm of duration of residence in months

$a_1 b_1, b_2$ are age-specific parameters,

was suggested.

As made clear by inspecting Land's (1969) results, however, a curvilinear relationship may be non-monotone and hence violate the duration-of-stay effect as expressed by the "Law of Cumulative Inertia." Hence, the pertinent relationship may be negative only for a certain range of values of duration.

* * *

Morrison, P.A. 1973. Theoretical issues in the design of population mobility models. Environment and Planning 5: 125-134.

A framework which integrates micro- and macro-model approaches is proposed for analyzing population mobility. This involves integrating concepts on how people perceive and respond to potentially advantageous moves, with an econometric-regression type model which allocates mobile individuals to centres.

Although noting such sources of variability in the decision to move as:

1) a threshold
2) life cycle
3) occupational constraints
4) prior migration experience,

the author's treatment of micro-level notions is limited to modelling the cumulative inertia effect (McGinnis, 1968) using a Markov renewal process. Then, the allocation of the flows of the mobile population (stayers ignored) is determined by pull effects, after Lowry (1966);

$$\ln M_j = ao + a_1 \ln U_j + a_2 \ln W_j + a_3 \ln C_j + a_4 \ln A_j + a_5 \ln D_{ij} ,$$

where M_j proportion of movers to j

U_j unemployment as % of total nonagricultural labour force

W_j hourly manufacturing wage

C_j labour force size

A_j number persons in armed forces

D_{ij} airline distance from i - j.

The proposals in this paper simply reiterate the suggestions in Ginsberg (1971, 1972) and Spilerman (1972a, 1972b).

* * *

Olsson, G. and S. Gale. 1968. Spatial theory and human behaviour. Papers, Regional Science Association 21: 229-242.

Suggest that some extended forms of the Markov chain model may be useful frameworks for studying spatial behaviour. They propose:

(1) augmenting the transition matrix (and hence distribution vector) by additional dimension(s), to allow for additional criterion variables; and

(2) altering the usual Markov transformation by including additional matrix operators.

The first suggestion has essentially been operationalized by McGinnis (1968), for one additional dimension. The second idea would render the model to be strictly non-Markovian. To retain the Markov framework, it may be possible to accommodate some of the weighting, non-linear, and local operator effects within the basic structure of the transition matrix, instead of introducing other forms of matrix operators and transformations. For example, to account for spatial contingency effects, some gravity expression can be used to estimate the transition matrix.

* * *

Rogers, A. and R. Miller. 1967. Estimating a matrix population growth operator from distributional time series. Annals, Association of American Geographers 57: 751-756.

A brief discussion of estimating the "growth operator" G in the "components-of-change population model":

$$W_{t+1} = (B-D+P')W_t = GW_t$$

where W_t population distribution vector at time t

B,D diagonal matrices denoting crude birth and death rates, respectively

P' migration transition matrix.

Estimation methods discussed were:

(1) unrestricted least-squares, which has a problem in that negative estimates, which are inadmissable within the model, are possible;

(2) minimum absolute deviations, which uses linear programming to determine the vector that minimizes the sum of the absolute deviations between observed and predicated values; aside from the assumption of a stationary G, a question arises in that absolute deviations, rather than some normalized criterion, are used.

* * *

Prais, S.J. 1955. Measuring social mobility. Journal of the Royal Statistical Society. Series A. 118:56-66.

From a random sample, transition matrix was computed in which the elements represented the probability of the son being in the j^{th} occupational class given that the father was in the i^{th} - so that the time interval was one generation. The actual distributions for fathers and sons were compared to the predicted equilibrium vector; all were appreciably different so that it was concluded that no equilibrium had been reached (actually only a comparison between the fathers and the sons was necessary to test this). Indices of mobility were suggested, one of which was,

$$M_j = (1 - P_j) / (1 - P_{jj}),$$

where P_j and P_{jj} are elements of the equilibrium vector and transition matrix respectively.

Also, it was suggested that the definitions of the states be allowed to change over time, with the likelihood of changes being represented by a transition matrix.

* * *

Stoyle, J. 1963. Land utilization as a stochastic process. Canadian Journal of Agricultural Economics 11.2: 52-64.

Applied Markov chain to analyzing the distribution of crop acreage over time. The factors influencing the sequence and number of acres planted to each crop were assumed to be government restrictions (resulting in a conservation reserve which was considered as an absorbing state), and the farmers' crop rotation patterns. Price, cost, new crop, and policy change factors could not be appreciably reflected within the stationary Markov chain. The transition matrix was estimated from four annual transition matrices derived from a sample from questionnaires. A Friedman two-way analysis of variance by ranks on selected yearly transition probabilities, was used to justify the claim that the yearly transition probabilities for given transitions were constant. The limiting vector, future distribution estimates, and first passage time properties were computed.

* * *

Rose, H.M. 1970. The structure of retail trade in a racially changing trade area. Geographical Analysis 2: 135-148.

Uses a simple Markov chain approach to study the effects of the spatial development of the Negro ghetto in major central cities on the retail structure of business clusters within its area. Notes that the influx of the Negro population changes the transition matrix of retail categories, resulting in a general decline in retail activity and increase in a few "Negro-related" activities.

A rather discursive analysis with no formal tests, such as χ^2, of transition stationarity; and no rigorous attempt to isolate any specific relationship between increased Negro population and retail change.

* * *

Smith, P.E. 1961. Markov chains, exchange matrices, and regional development. Journal of Regional Science 3: 27-36.

Investigates the possibility of increasing the incomes of some regions relative to those of others; income gains are achieved through transaction propensities, which are expressed in terms of Markov chains. To counter the sustained disparities which may be characteristic of an equilibrium situation, three solutions are suggested:

(1) in a reducible chain, regions can be aggregated into irreducible submatrices; a one-time injection into the economy of one of these submatrices will then be sustained and increase income relative to the other irreducible submatrices;

(2) however, for one region within a set of regions in an irreducible matrix continuous injections are required to offset the assymptotic convergence to the equilibrium state of regional disparities;

(3) alternatively, non-stationarity must be allowed, so that fiscal and monetary policies may alter propensities to trade, and hence alter the transition matrix over time.

* * *

Spilerman, S. 1972. Extensions of the mover-stayer model. American Journal of Sociology 78: 599-627.

Develops a solution to an extension of the mover-stayer model. The proposed extension is based on different rates of mobility in a heterogenous population, assuming that individual's transitions follow a Poisson process. Now, if individual rates of mobility are assumed to follow a gamma density, a negative binomial will describe the population's movement frequency distribution:

$$r_\nu(t) = \binom{\alpha + \nu - 1}{\nu} \left(\frac{t}{\beta+t}\right)^{\nu} \left(\frac{\beta}{\beta+t}\right)^{\alpha}$$

where $r_\nu(t)$ expected proportion of total population making transitions in time t;

α, β parameters of gamma distribution.

Although each individual is allowed a different rate of mobility, the same individual-level transition matrix, M, is used. Hence, the overall one-step transition matrix is:

$$P(1) = \sum_{\nu=o}^{\infty} r_\nu(1)\, M^\nu$$

$$= \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left[I - \left(\frac{1}{\beta+1}\right) M\right]^{-\alpha}$$

Empirical estimates of α, β and one of the matrix variables M or P(t) allows estimation of the other matrix variable.

A migration empirical example revealed some operational problems:

(1) in the analysis, people with greater than ν=4 moves were omitted, eliminating the extremely mobile portion of the population;

(2) little interregional migration since only 4 states were used; such aggregation limits the information gained from the analysis;

(3) underestimation of moves in cases where asking about the "distant" past;

(4) stationarity of M violated, due to life cycle changes.

* * *

Spilerman, S. 1972b. The analysis of mobility processes by the introduction of independent variables into a Markov chain. American Sociological Review 37: 277-294.

As an alternate to disaggregation, Spilerman treats the problem of population heterogeneity via regression, which allows a Markovian framework, but with the advantages of considering: (i) sources of variation; (ii) transition matrix changes due either to population shifts, or actual changes in the process; (iii) individual level matrices.

The regression model will be briefly outlined:

(1) m^2 regression equations are constructed, where m is the no. of states (i.e. one equation corresponding to each element of the transition matrix); the dependent variable

$$Y_{ij} = \begin{cases} 1, & \text{if a transition from i to j occurs in 1 step} \\ 0, & \text{otherwise ;} \end{cases}$$

the independent variables reflect past residential location and socio-economic attributes.

(2) Summary over all individuals gives n_{ij}, the no. moving from i to j; and transition probability is then $P_{ij} = n_{ij}/n$.

(3) Reestimate the Y_{ij}'s at each subsequent time interval, to allow for changes in the process.

(4) Although the essential process itself may not change, the feedback effects of changes in population distribution may in themselves alter the P_{ij}'s. To investigate the significance of these effects, $\hat{P}_1(1)$ and $P_o 1$ are compared to $P_1(1)$ - where $P_1(1)$ is the actual one-step transition matrix using regression coefficients from interval 0 but data on independent variables from interval 1. If $\hat{P}_1(1)$ is significantly more similar than $P_o(1)$ when compared with $P_1(1)$, and is in fact very similar to $P_1(1)$, then one may conclude that changes in transition probabilities were not due to changes in the process (i.e. in the regression coefficients), but to demographic changes reflected in the values of the variables.

Several issues are now discussed:

(a) First, it can be noted that although Spilerman contrasts this model to a previous one (Spilerman, 1972a) characterized by variable mobility rates but fixed transition probabilities, the present one implicitly addresses the factor of mobility variation via the duration of residence (variables X_{11} and X_{14}).

(b) The Y_{ij} should not be binary, since it should represent the probability of moving for an individual. What results in the binary situation corresponds to the unnecessarily restrictive movers-stayers dichotomy. Of course, for purposes of operationalizing the analysis, a binary measure was necessary since one does not know what preferences are attached to different transition possibilities, just what transition is actually made.

(c) The R^2 values were reported to be low ($<.05$), but no F-ratio's were given; it is the latter which is often more informative. Also, no t-statistic values were given for the regression coefficients.

(d) Although it was suggested that the approach differs from McFarland (1970), which aggregates individual matrices, to some extent this appears to

be done here as well. The Y_{ij}'s were summed for all individuals.

(e) For forecasting, the approach is somewhat restricted in that forecasts of the independent variables would be required, even for constant regression coefficients. And if the latter changed over time, it would appear that little information would be gained over what would already be known.

* * *

Stone, R. 1972. A Markovian education model and other examples linking social behaviour to the economy. Journal of the Royal Statistical Society, A 135: 511-543.

Presents a matrix framework for analyzing human stocks and flows, and for relating them to economic variables. However, little is done to explicitly integrate these components within one system, other than to note the structural correspondence between a Markov matrix and an input-output matrix.

The treatment of allowing transition probabilities to vary with the past transitional history (p. 516) simply appears to reduce to a case of the cohort-survival class of models which have been discussed in some detail by Rogers (1968). Further, the treatment of a non-stationary transition matrix (p. 517-518) presents two additional issues. First, the formulation imputs an iterative operator on the transition matrix C, denoted by Λ , which is itself assumed stationary. And second, the fact that transient behaviour is allowed may preclude the general existence of an equilibrium; hence the consideration of the fundamental matrix, analogous to the stationary case, appears extraneous.

Several examples given in which the states are education-related attributes.

* * *

Sykes, Z.M. 1969. Some stochastic versions of the matrix model for population dynamics. Journal of the American Statistical Association 64: 111-130.

Presents three stochastic versions of the deterministic population dynamics model,

$$X_{t+1} = A_t X_t , \qquad [1]$$

where

X_t is distribution vector at time t

A_t transition matrix at time t.

First, [1] is subject to additive (independent) random errors ε_t, with mean $E(\varepsilon_t) = 0$ and covariance $\mathrm{Cov}(\varepsilon_s, \varepsilon_t) = \Gamma_{st}$. Three main problems were noted:

i) errors are likely to be correlated for small $\Delta t = (t+1) - t$;

ii) negative estimates of χ_{t+1} are possible;

iii) as t increases, it may be shown that either the mean of χ_t increases without formal, or χ_t tend to a constant plus a moving process (essentially removing the multiplicative nature of the growth process).

Second, the elements of the transition matrix A_t represent probabilities, rather than rates. Briefly, it is assumed that:

i) each member of the i^{th} group (state) at time t has probability S_i of surviving to be a member of the (i+1)st group at time t+1;

ii) b_i is the probability of a member in i of contributing (giving birth to) a member, who would then be in the first group;

iii) all events result from mutually independent binomial trials.

Three points may be raised:

(a) the model is applicable only for a certain class of processes, such as age distribution of a population; this is because no allowance is made for transitions to any "previous" state or any state other than the "next" one;

(b) the assumption of independent events would be a major shortcoming when dealing with any process involving political, economic, or spatial contingencies;

(c) the variance of the estimates decreased the larger the population.

Third, the transition matrices $\{A_t\}$ are random variables:

$$\chi_{t+1} = (A_t + \Delta_t)\,\chi_t$$

where $\{\Delta_t\}$ is a sequence of independent matrix random variables with

$$E(\Delta_t) = 0 \text{ and } \mathrm{Cov}(\Delta_i, \Delta_j) = \Sigma_t.$$

Again, as in the first model, the Δ_t would only be uncorrelated for sufficiently large Δ_t.

Further, although not discussed by Sykes, it seems possible that extreme values of Δ_t may generate negative estimates of χ_{t+1}.

Expressions were computed for the first and second moments in each version. As expected, the mean corresponded to the deterministic process itself.

* * *

Tarver, J.D. and W.R. Gurley. 1965. A stochastic analysis of geographic mobility and population projections of the census divisions in the United States. Demography 2: 134-139.

Transition matrix computed from 1955 to 1960 data used to project estimated population distribution in 1965 and 1970 for the nine census divisions in the United States.

* * *

White, H.C. 1971. Multipliers, vacancy chains, and filtering in housing. American Institute of Planners, Journal 37: 88-94.

The act of a household leaving a housing structure is considered to be a vacancy. Subsequent moves by households into and out of this structure form a chain, whose average length is the multiplier, the total number of moves caused by the initial arrival of a vacancy. It is suggested that properties of chains and the multiplier should serve as a basis for policy choices. A hypothetical example was used to compute the multiplier from a tabulation of the percentage of all vacancy chains which are of length j. This latter value was given as:

$$F\,Q^{j-1}\,p$$

where F is a row vector giving probabilities of a new house or one whose occupants have left the system being in a certain price range.

Q a square matrix giving the probability of a move from a dwelling in price range i to a vacancy in price range j.

p a column vector giving the probabilities of transitions from price range i to the dwelling being demolished or occupied by a household from outside the system.

It can be noted that emigration and immigration were treated in a somewhat unusual manner, being equated to housing construction and demolition respectively. Also, it would have been more straightforward to aggregate F, Q, and p (and adding another element and redefining F) to give a matrix of the form:

$$\left(\begin{array}{c|c} Q & p \\ \hline F' & x \end{array}\right)$$

which can be used to calculate the mean absorption time, which is simply what White calls the "multiplier."

* * *

Wolfe, H.B. 1967. Models for condition aging of residential structures. American Institute of Planners, Journal 33: 192-196.

Constructs transition matrices for changes in the condition of housing over a six year period. Different matrices were computed for different types and ages of housing structures. It was suggested that different matrices could be constructed under varying conditions (policies), such as a conservation program, increased code enforcement, a rehabilitation program, and so on. But at the same time, it was noted that the corresponding data requirements would be rather severe.

APPENDIX B

Part II: Economic Models Pertaining to Urban Growth

R. John Miron

1. General Economic Models

Kain, J.F. and J.R. Meyer. 1968. Computer simulations, physio-economic systems, and intraregional models. American Economic Review 58: 171-181.

Kain and Meyer review the development of regional economic models, the similarities among different kinds of models, and the general weaknesses of then-current models.

They isolate three major deficiencies in intraregional or metropolitan growth models; (i) the inability to adequately model industry location, (ii) the inability to handle the housing market in regard to the issue of discrimination, and (iii) the inability to understand the economics of housing.

Since many current land-use and economic growth models use the concepts of basic and nonbasic employment and since most also presume that basic employment is exogenously given, it is difficult, they argue, to justify the lack of knowledge on the determinants of industry location.

They argue that there is also a bias toward overestimating employment levels within central city areas because of a failure to model the adjustment process by which the stock of structures is adapted to new uses over time.

Finally, they assert that housing market segregation has produced special economic phenomenon which has not been incorporated into urban growth models.

Since their paper was written, a number of models have been

constructed which attempt to handle some of these conditions. It is not clear however that these have been entirely adequate and any new intraurban model should attempt to stress these areas.

2. Intraurban Economic Models

Barr, J.L. 1972. City size, land rent, and the supply of public goods. Regional and Urban Economics 2:67-103; Koenker, R. 1972. An empirical note on the elasticity of substitution between land and capital in a monocentric housing market. Journal of Regional Science 12: 299-305; Mills, E.S. 1967. An aggregative model of resource allocation in a metropolitan area. American Economic Review 57: 197-210; Muth, R.F. 1971. The derived demand for urban residential land. Urban Studies 8; Niedercorn, J.H. 1971. A negative exponential model of urban landuse densities and its implications for metropolitan development. Journal of Regional Science 11: 317-326; Wingo, L. 1961. Transportation and Urban Land. Washington: Resources for the Future.

Since Von Thunen introduced his theory of agricultural location, urban researchers have attempted to construct analogous models of household and commercial location. Alonso (1960) and Wingo (1961) seem to have set the framework for current models of this type as can be seen from the later work of Mills (1967, 1972), Muth (1969, 1971), Barr (1971), Koenker (1972) and Niedercorn (1971).

Let us concentrate on Mills' 1967 model as an example. Mills presents an initial version of his model of urban land allocation in this paper; later versions being presented in his 1972 book. This model consists of a set of assertions about the production of and demand for three goods; housing transportation, and a composite good. In this model, he assumes a concentric, circular CBD in which production of the composite good takes place and a surrounding 'suburb' in which all housing production takes place. Transport

goods are produced both in the suburb and in the CBD. Production draws on three factors: land, labour, and capital. Using these assertions and some assumptions about equilibrium behaviour, Mills is able to derive the efficient allocation of the factor land among the three output goods. Within this overview, Mills' model may now be drawn out in more detail.

(a) Production. He allows for a Cobb-Douglas production function for the composite good (X_1) and the housing good (X_3) with constant-returns in the latter. He presumes a proportional relationship between transport output and land input for the remaining good so that he effectively deals with a single transport mode or good.

(b) Demand. The composite good is assumed to face a market demand of finite constant elasticity while the housing market is perfectly competitive. Further, housing demand per resident is assumed to be independent of location. Road demand at a given radius in the suburbs is assumed to be proportional to the number of workers who reside further from the CBD centre and who commute to the CBD. In the CBD, road demand at a given radius is proportional to the number of workers employed closer to the centre. The price of the transport good is assumed to be proportional to the rental rate of land.

(c) Equilibrium. Mills uses the popular housing price-equilibrium condition that the change in housing prices must be offset by a corresponding change in transportation costs. Land rent at the city fringe must further equal the agricultural land rent and rent at the CBD fringe must be equal for housing and composite good production. Finally, all land in the city must be allocated to some use and all workers must be employed and live somewhere.

Using these three sets of conditions, Mills is able to derive

(i) the equilibrium allocation of land among the three products, (ii) the equilibrium land-rent gradient, and (iii) the equilibrium residential density.

Barr (1972) reworks Mills' basic model in a number of directions. He initially establishes a model in which there is one point of production and all other land is devoted to housing. He is able then to derive land rent and density functions for a particular class of utility-maximizers in a simpler form than in Mills' model. Barr expands this model to include a finite-sized area CBD with a particular class of profit-maximizing producers as Mills has done. By ignoring transport goods, Barr is able again to come up with a simpler solution.

Barr then steps beyond Mills' analysis and examines the impact of pure public goods on this model. He introduces the public good into the utility functions of residents. In one model, the public good is financed through proportional land-value taxation. In another, he allows for housing production and has the public good financed through proportional site-value (land and building) taxation. Barr allows for a fixed head tax in a third model. In a final model, he considers the case in which a monopsonist is permitted to determine the amount of the public good and to set a corresponding tax rate on land values.

In his conclusion, Barr offers a number of suggestions about areas of future research. He suggests modifying his model to allow for a distribution of wages. He further suggests that the model also be modified to allow for economies-of-scale in production. Finally, he asserts that the urban growth process which could be modelled here would have important growth-limiting aspects. In particular, he sees that intraurban access will limit the response of labour in a region to a wage differential at a particular urban centre.

Muth (1969) has refined a monocentric location and production theory of housing. In this work, extensive use is made of the Cobb-Douglas production function in regard to housing production. As is well-known, the C-D function implicitly assumes that the elasticity of substitution among factor inputs is unity. In other words, if the production of housing is achieved using the two factors land and capital, there will be a one percent increase in the ratio of land to capital if the price of capital increases one percent relative to the price of land.

Koenker (1972) argues that this constant production elasticity is not in fact unity. He uses a Constant Elasticity of Substitution production function together with the usual assumptions about a competitive, profit-maximizing housing market to derive an equation suitable for the estimation of this elasticity. Koenker uses data on multi-family housing in Ann Arbor, Michigan, to estimate this elasticity at 0.71. Assuming an exponential decay in land prices from the centre of Ann Arbor, Koenker derives an expression for housing prices based on this land price gradient, the rate of factor substitution, and distance from the city centre.

Muth (1971) undertook a similar study of the rate of factor substitution. Using an equivalent model applied to data on new FHA-insured single-family-detached houses in 1966, Muth finds this elasticity to be around 0.50.

This debate over the elasticity of substitution may, at first glance, appear to be quite esoteric. There is an implication however which is quite important. The higher the rate of substitution, the more builders are able to substitute capital for land. This means that the housing supply price gradient over space will be more gentle than the land rent gradient. Changes in the rate of substitution over time thus have an effect on the

residential density gradient. In this way, changes in residential density can be related not only to the more widely accepted notion of changes in access but also to changes in housing technology.

Niedercorn's contribution has been to show that under fairly general conditions, the Von Thunen model applied to urban areas results in a negative exponential density gradient. It is noteworthy that he does this using a C-D production function since, as has been noted above, permitting a non-unity elasticity of substitution may serve to make the density gradient steeper or more gentle.

With all of these refinements and adaptations of the basic Von Thunen model, there continues to remain a stubborn equilibrium problem. In all these models, land is allocated to the highest bidders. But, these bids are made on the basis of some set of prices somewhere else in the system. Furthermore these prices are not entirely independent of the way in which land is allocated. As an example, it is easily seen in all these models that the supply of population and workers is assumed to be infinitely elastic. Population moves in and out of the urban region with every change in parameter values. Is this realistic in any sense? Can we expect production or population which is priced out of a location in the city to respond by moving out in the short or even the long-run?

These models are useful exercises as far as they go, but in respect of this equilibrium condition at least they have not progressed beyond Alonso.

Engle, R.F., et al. 1972. An econometric simulation model of intra-metropolitan housing location: housing, business, transportation, and local government. American Economic Review 62.

Engle, et al are currently developing a model of the Boston SMSA consisting of three main sections; a macroeconomic non-spatial model of output, employment, and income distribution, a model of long-term stock adjustments, and a model of spatial allocation.

As a policy-oriented model, they see three requirements for their model; (i) to generate policy-relevant indicator variables, (ii) to provide variables by which policies can be represented, and (iii) to provide a model with a high degree of predictive power.

Into the macroeconomic model are placed population, capital stock, technology, external demand, wages, prices, and unemployment. Outputs include local employment, wages, prices, and an income distribution.

The stock adjustment model is designed to relate stock changes to conditions both within Boston and within the rest of the nation. This model, in conjunction with the macroeconomic model, provide an aggregate growth model for Boston.

The outputs of these two models enter as inputs into the spatial allocation model. This model allocates location patterns within the city and may lead to feedbacks to the other two models through housing and land prices and average tax rates.

While appearing to be quite complicated, these models are actually quite simple in terms of theoretical structure. A great deal of complication arises simply from the very disaggregate form of these models, particularly the spatial allocation model.

Moses, L. and H.F. Williamson. 1967. The location of economic activities in cities. American Economic Review 57: 211-222.

Moses and Williamson propose a simple theory to explain the emergence of the core-dominated commercial city in the nineteenth century and its relative decline in the twentieth. They argue that transportation technology is at the root of both trends.

In the nineteenth century, it is argued that the emergence of the railroad caused economic activity to become core-dominated. The cost of intra-urban goods movement was very high so that all goods-producing activity had to take place in the CBD. Workers with relatively low commuting costs were allocated to the non-CBD periphery.

They argue that the emergence of truck transport in the twentieth century changed this. In the first phase of this change, trucks emerged primarily as intra-urban carriers. This allowed firms to seek locations outside the CBD although there were still tied to the CBD in that inter-urban activity still originated there. In the latter phase of this change, the tie to the CBD was completely broken with the development of interurban trucking.

This argument is in accord with those proposed by Mills (1972) and Barr (1972) who argue that a sufficient condition for a core-dominated city is that the marginal cost of shipping per unit distance from the centre increase more quickly than the marginal cost of commuting.

Paelinck, J. 1972. Formal allocation models for urban activities. Review of Regional Studies 2.

Paelinck begins by noting the large literature on consumer expenditures in economics. In one of the more common models in this literature, consumer expenditures on good 'i' (P_iQ_i) are related to the consumer's total budget (B) and to the price of good 'i' (P_i). The form is as follows:

$$P_iQ_i = \gamma_i (P_i/B)^{\delta_i} B/ \left(\sum_{j=1}^{N} \gamma_j (P_j/B)^{\delta_j} \right)$$

Paelinck asks why such a model might not be used to replicate the land-purchasing behaviour of firms. He suggests that a major problem might be the definition of a suitable price measure and offers the notion of density (D_i) as a proxy for land prices. By density, he means employment per unit land. The quantity of the good land purchased by industry 'i' (L_i) is now the good consumed. The total 'budget constraint' is now given by

$$\sum_{j=1}^{N} D_j L_j$$

which is nothing more than aggregate employment (E). We may thus rewrite the above allocation problem as the following.

$$D_iL_i = \gamma_i(D_i/E)^{\delta_i} E/\left(\sum_{j=1}^{N} \gamma_j (D_j/E)^{\delta_j} \right)$$

Paelinck's model certainly looks interesting as an allocation method. It is not clear yet, however, just how critical the use of the density proxy for price is to this model and how justifiable it is.

3. Urban Aggregate Models

Borcherding, D.R. and R.T. Deacon. 1972. The demand for the services of non-federal governments: An econometric approach to collective choice. American Economic Review 62: 891-901.

Borcherding and Deacon formulate a model of aggregate public spending at the local level. In this model, public decision-making is presumed to be related to the interests of the median voter.

Production of public services is assumed to be efficient and is related to the two factor inputs, labour and capital, via a constant returns Cobb Douglas function. Assuming at any point in time that the rate of return on capital is fixed, it is shown that the preceding implies that the marginal cost of production of services, C_x, is log-linearly related to the wage rate, W.

Consumption of services, Q, by the individual median voter is related to the aggregate output of services, X, and the aggregate number of voters, N.

$$Q = X/N^{\gamma}$$

If γ is zero, then the service is a pure public good with everyone consuming the same total output. If γ is unity, the service is a pure private good with everyone consuming an equal share of the output (or at least every voter near the median getting society's average). Presumably γ lies between zero and unity for any service. Since the median voter is assumed to share equally in the tax load, his share of the cost of production is C_xX/N and his subjective price for this consumption is

$$S = (C_xX)/(NQ)$$

Finally, assume a log-linear demand function for these public services,

by the median voter, as a function of his income, Y, and the subjective price of public services, S.

Using these assumptions about the production of, distribution of, and demand for public services, Borcherding and Deacon are able to derive a log-linear relationship between aggregate expenditures and (i) the wage level, (ii) the population size, and (iii) the median income. They estimate this model using aggregate state data for the US with a breakdown into eight expenditure functions.

A major problem encountered is the persistently-high estimates obtained for γ. Typically, this estimate is greater than unity and suggests that the median voters are successfully distriminating against other voters. The authors are not presently able to suggest how such a discrimination relationship might be introduced into the model.

This model seems to be a fairly-elegant yet simple method for introducing a public sector into a dynamic model of urban change. Their problems with γ estimates suggest, however, that some care need be taken in any application of the model.

Harrison, D. and J.F. Kain. 1970. An historical model of urban form. Paper prepared for the CUE Research Conference, Chicago, Illinois, September 11-12, 1970.

There seem to be some regularly recurrent themes in urban research. Wingo (1961), for instance, has argued that urban models and theory should be concerned with understanding the changes in urban structure rather than the whole structure itself. The argument is that there are too many historical differences among cities to make stock models useful or even

usable generally. Models which emphasize changes in stocks, i.e., flow models, are much more likely to be useful because they tend to minimize the influence of historical factors.

Harrison and Kain revive this notion again in looking for a simple model to explain the ratio of single-family-detached to total dwellings. Instead of examining the stock form of the ratio, they choose the flow form and find that they are able to get reasonable results by relating the ratio of change in single-family-detached to change in total dwellings (S_t) to the time period (t) and the total dwelling units at the outset of the time period (H_t).

$$S_t = f(t, H_t)$$

This emphasis on incremental performance over time appears to be quite useful. In simple dynamic models of urban structure, this emphasis is not only useful but necessary.

Kain, J.F. 1967. Postwar metropolitan development: housing preferences and auto ownership. American Economic Review 57: 223-234.

Kain considers the question of the causal relationship between housing preference and auto ownership. Some people have forcefully argued that it is widespread automobile ownership which has shaped urban areas in the last fifty years and determined the spatial distribution and mix of housing types. Transportation planners have insisted however that the spatial distribution of housing, typically measured by net residential density, determines the level of auto ownership. Kain undertakes a test, using data on 54 municipalities in the Boston Metropolitan Region, in which he allows for both causal

relationships and for an interdependent relationship between these variables. His models are given by the following pair of equations.

$$D^* = \gamma_0 + \gamma_1 X_1^* + \gamma_2 X_2^* + \gamma_3 X_3^* + \gamma_4 X_4^* + \gamma_5 A^*$$

$$A^* = \delta_0 + \delta_1 X_1^* + \delta_2 X_2^* + \delta_3 X_3^* + \delta_4 X_5^* + \delta_5 D^*$$

In the first equation, log Net Residential Density (D^*) is related to median family income (X_1^*) logged, log median family size (X_2^*), log ratio of autos owned by households to total autos (X_3^*), the logged labour force participation rate (X_4^*), and logged rate of auto-ownership (A^*). In the second, A^* is related to X_1^*, X_2^*, and X_3^* as well as the log ratio of employment to households (X_5^*) and log density (D^*).

His findings are mixed and Kain suggests that he has not been able to satisfactorily answer his question. It is interesting to note though that the derived elasticity of D with respect to A is, on average, about ten times larger in absolute value than the elasticity of A with respect to D.

Kain's hypothesis would seem to be a useful one to keep in mind when constructing a larger scale urban model. This interrelationship certainly seems to be one which is often ignored unnecessarily. It would also be useful to rework this model to give it some theoretical underpinnings to tie it more closely to the main body of economics.

Oates, W.E., E. P. Howrey and W.J. Baumol. 1971. The analysis of public policy in dynamic urban models. Journal of Political Economy 79: 142-153; Baumol, W.J. 1972. The dynamics of urban problems and its policy implications. In Essays in Honour of Lord Robbins. M. Preston and B. Corry (eds.). London: Weidenfeld and Nicholson. 380-393.

Oates constructs a simple dynamic model of the growth of income in a central city. The dynamics of the middle-class flight to the suburbs is examined here. Percapita income at time 't', Y_t, related to the deterioration of the central city at the time 't-1', D_{t-1}. Current deterioration is related to current percapita income and to current government expenditures, E_t.

$$Y_t = a - gD_{t-1}$$

$$D_t = c - dY_t - eE_t$$

Finally the responsiveness of income to changes in immediate past deterioration, g, is related to the level of municipal taxes, T_t.

$$g = b + fT_t$$

This system of equations represents a dynamic process. A first order difference equation in income may be derived whose dynamic properties are well-known.

$$Y_t = (a-(b+fT_t)(c-eE_{t-1})) + d(b+fT_t)Y_{t-1}$$

Oates emphasizes that, in this dynamic model, the policy variables are taxes, T, and expenditures, E. Making income, Y, a policy variable is not useful in the long run because income will always tend toward an equilibrium defined by T and E.

Baumol (1972) repeats this model and then extends it somewhat. Middle class urbanites, M_t, are made up of those residing in the central city, M_{ut}, and those residing in the suburbs, M_{st}. Deterioration in the central city, D,

is negatively related to M_{ut} in his first hypothesis. Crowding in the suburbs, C_t, is asserted to be positively related to M_{st}. Now, M_{ut} is linearly related to D_t (negatively) and to C_t (positively). These hypotheses are shown to result in a first order difference equation in M_{ut} similar to the one derived above.

Baumol then introduces a population of poor persons, P_t, split again into central city, P_{ut}, and now rural, P_{rt}, populations. Deterioration of the central city is now directly related to P_{ut} and inversely related to M_{ut}. The number of urban poor is directly related to the previous rural poor and inversely related to the number of urban poor.

These two sub-models can be represented as follows.

$$M_{ut} = \gamma_0 + \gamma_1 M_{ut-1} + \gamma_2 P_{ut-1}$$

$$P_{ut} = \delta_0 + \delta_1 P_{ut-1}$$

This model describes the flows in and out of the central city by poor and middle class persons; the poor arriving from rural areas and the middle class in flight to the suburbs. Baumol notes the similarity between this model and Volterra models.

4. Interregional Economic Models

Borts, G.H. 1968. Growth and capital movements among U.S. regions in the postwar period. American Economic Review 58: 155-161.

Borts presents a model to analyze borrowing and lending among regions. In this model, net export flows of commodities must be matched by corresponding capital flows which, in turn, are the excess of local savings over local investment. Under the assumptions of full employment and a competitive equilibrium, Borts is able to show that the ratio of regional

received income to regional produced income becomes constant over time. In other words, the flow of income on external investment becomes constant relative to regional produced income.

It is difficult to see, at the outset, the relevance of Bort's model to urban dynamics. What Borts does make quite clear, however, is that exports and capital growth must be related. This is not a particularly new idea to international economists but is definitely missing from general regional analysis.

Jutila, S.K. 1971. A linear model for agglomeration, diffusion, and growth of regional economic activity. Regional Science Perspectives 1:83-108.

Jutila defines a simple Keynesian multiplier-accelerator model of growth to a continuous production plane. At each point on the plane, the identity holds that

$$Y(x,t) = C(x,t) + I(x,t) + E(x,t) - M(x,t)$$

where total income density (Y) is related to consumption density (C), investment density (I), export density (E), and import density (M) at each point in space, x, and in time, t.

The multiplier effect arises from the hypothesis that consumption is proportional to income density and the accelerator effect arises from the hypothesis that investment is proportional to the rate of change of income or output over time.

$$C(x,t) = bY(x,t)$$

$$I(x,t) = a(dY(x,t)/dt)$$

By substituting these two hypotheses into the earlier identity, a differential

equation can be obtained.

$$(1-b)Y(x,t) - a(dY(x,t)/dt) = E(x,t) - M(x,t)$$

Jutila derives one final hypothesis about the net level of exports. For reasons which are not made clear in his paper, Jutila hypothesizes that the level of net exports is proportional to the rate of rate of change of production over space.

$$E(x,t) - M(x,t) = c(\partial^2 Y(x,t)/\partial x^2)$$

Combining this hypothesis into the preceding equation yields an expression solely in terms of income, Y.

$$(1-b)Y(x,t) - a(dY(x,t)/dt) - c(\partial^2 Y(x,t)/\partial x^2) = 0$$

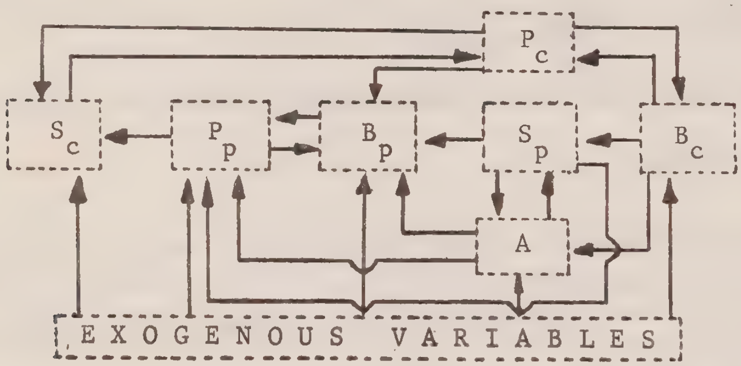
Jutila is then able to solve this last equation using Laplace transforms and derives the response function.

As Jutila admits in a later paper, this is a short-run demand-oriented model not unlike the much criticized export-base model. There are virtually no supply conditions in the model so that resource constraints such as labour shortages cannot be taken into account. More surprisingly, there is no attention paid to capital imports and their determinants. This would imply that even capital resource behaviour has not been satisfactorily considered in the model.

Lewis, W.C. and J.R. Prescott. 1972. Urban regional development and growth centres: an econometric study. Journal of Regional Science 12:57-70.

Lewis and Prescott consider the relationship between medium sized central cities and their hinterlands. In particular, they examine FEA's in the United States to measure the relationship between the growth in

employment and population of the central city and similar growth in the remainder of the FEA. Their dependent variables include manufacturing employment (B), service employment (S), agricultural employment (A), and population (P). All variables except A are defined separately for the central city (c) and the peripheral FEA (p). The structure of their model is schematized below and is complicated but static.



They present almost no justification for this interdependent structure. There would have to be some sort of theoretical basis given to this model before one could feel very easy about using it.

5. Regional Aggregate Models

Bebee, E. 1972. Regional housing markets and population flows in Canada: 1956-1967. Canadian Journal of Economics 5:3:386-397.

Bebee establishes a stock model of housing in which, in reduced form, he asserts that housing starts (HS) are a function of net in-migration of families (PFLOW), average mortgage rates (MTGE), the spread between mortgage and industrial bond rates (CREDIT), the volume of CMHC-approved mortgages (CMHC), and per-family disposable incomes (FAMYD). He estimates

this model for Canada and for five sub-regions (Atlantic, Quebec, Ontario, West, BC). In each of these applications, PFLOW and CREDIT are hypothesized to have distributed-lag effects of up to twelve quarters on housing starts.

Bebee estimates the impact on housing starts of an increase of 100 family units of net in-migration. Nationally, this impact is 82 units over 8 quarters. In the outflow regions (Atlantic, West), it is lower at 22 to 42 units and concentrated in 5 to 6 quarters. The net inflow regions have impacts of 24 to 51 units spread over 8 to 11 quarters. He concludes that there are significant regional differences in the response of housing markets to changes in in-migration and credit conditions although he offers no explanation for the cause of these differences.

This paper provides a useful framework for thinking about the effect of in-migration on a city's housing stock. There are interesting lag-effects here which would serve to render a growth model more dynamic.

Casetti, E. n.d. On the sensitivity of regional incomes to export fluctuations. Discussion Paper No.2. Department of Geography. The Ohio State University.

Casetti forms a model of regional income fluctuations to separate out local fluctuations from variations in export demand. He begins with the familiar multiplier-accelerator model linking regional income (Y) to regional consumption (C), regional investment (I), and regional exports (E) at time 't'.

$$Y_t = C_t + I_t + E_t$$

$$C_t = a + bY_{t-1}$$

$$I_t = c + d(Y_{t-1} - Y_{t-2})$$

He reduces the system to a second order difference equation in Y and solves this to get income as a function of time and exports. This model may have fluctuations even when E is held constant and these variations are affected by whatever oscillations E may itself bring. Casetti develops equations to show the sensitivity of the response of Y to a fluctuating E under varying assumptions about the magnitudes of the parameters.

What Casetti has done here is to provide us with a model for relating short-run cycles in regional activity to changes in national activity. As he has shown, it may be possible for quite difference cycles at the regional and national levels to exist and be consistent with one another.

* * *

Dagenais, M. 1973. Un modele annuel de prevision pour l'economie du-Quebec. Canadian Journal of Economics 6:1:62-78.

In the following paper, two regional econometric growth models are discussed: that of Glickman and that of Czamanski. Dagenais' model is similar to these.

Dagenais presents an annual economic growth model for the Province of Quebec. This model contains 18 stochastic and 5 identity equations. This model is based on a capital-shift hypothesis where it is assumed that population growth (i.e., net in-migration) is independent of economic growth. Labour force can and does adjust in the short-run to changes in economic conditions but is presumed to be fixed in a long-run trend relationship to exogenous population growth. Employment is further seen to be related to intra and extra-regional aggregate demand as well as to labour force and population.

$$\begin{tikzcd}
\text{POPULATION} \arrow[r] & \text{LABOUR FORCE} \arrow[d, shift left] \\
& \text{EMPLOYMENT} \arrow[u, shift left] \arrow[d, shift left] \\
& \text{AGGREGATED DEMAND} \arrow[u, shift left]
\end{tikzcd}$$

Dagenais' assumption of an exogenous population variable is somewhat unsatisfactory for a regional model. At the scale of an urban model, it becomes even less acceptable because it ignores the importance of in-migration fluctuations.

Golladay, F.L. and A.D. Sandalov. 1972. Optimal development policy in an open regional economy: A programming analysis. Journal of Regional Science 12: 185-198.

Golladay and Sandalov derive a regional economic planning model for the state of New Mexico. In this model, an input-output model of twenty sectors is used in conjunction with constraints on capital, labour, savings, interregional flows, private consumption, state purchases, exports, production, and investment. An objective function in which output is maximized is set up. The government is seen to have policy instruments related to current and investment expenditures for each industry sector. They show how the objective function and constraints are linear and thus how the problem resolves to one of formulating an appropriate linear programming problem.

One should note that (i) there are few behavioural conditions in this model, (ii) the model is dynamic only over a single time period (no multi-period optimization), and (iii) the government has only one objective, the maximization of output. Consequently, this model is of limited usefulness in its present form.

URBAN SYSTEMS AND ECONOMIC GROWTH: A REVIEW

JOHN MIRON

Much attention of urban researchers has been focussed on the study of interurban systems or groupings of interacting cities.[1] A good deal of this literature has been more descriptive than analytic and more conceptual than operational. This focus on description and concept has, perhaps, been justifiable in view of increasingly complex hypotheses as to how interactions might take place, hypotheses which quickly digress from traditional economic theory into social, political, and ecological considerations. However, relatively simple analytic models are also quite valuable to researchers and policy-makers and these have been unduly neglected.

At the same time, the application of economic theory to the development of urban growth models has also been deficient. In the main, such models have attempted to relate urban growth to factors which are inherent to the urban centre being studied or to its relative share of regional or national growth. Very little comprehensive modelling has been done, outside of the crude and preliminary work initiated by Lösch and Christaller, on the extent to which the growth of an urban centre is dependent on the growth of nearby urban centres.

---

[1]Although the emphasis in this paper is on interurban systems, it is clear that there are a great many similarities between these and intra-urban systems. Thus, the findings of this paper should be generally if indirectly applicable to the study of intraurban systems.

There appears, therefore, to be considerable room for the development of a growth model which relates growth at each urban centre in the region or system to a number of factors inherent to those centres as well as the characteristics of the networks linking the centres within the system. Networks, as used here, include both the physical structures and the activity patterns which take place within them. This is not to say that such models have never been developed before since several complicated, if theoretically deficient and inelegant models have been estimated already.[2] Where research appears to have been inadequate is in the development of simple, numerically estimable models which are theoretically robust.

1. Framework for an Urban Economic Growth Model

It would appear to be not too difficult to formulate a simple analytic economic model of an interurban system, at least at an aggregate level. Such a model might be partitioned into two sections, the states of the various cities and the interactions between the states of various pairs of cities. A state is thought of here as a set of elements, made up of relationships and variables, which are internal to the city and which determine its economic well-being. An interaction on the other hand is a relationship among elements of the states of a pair of cities.

In an aggregate model, one might designate a relatively simple set of elements as the state of a city. Among the variables included as elements might be output, factor inputs, technology, prices, and incomes. Among the relationships included as elements might be those linking factor

---

[2] Among the most ambitious of these is the massive study by Roberts and Kresge (1968) an Columbia.

inputs, technology, and output to factor prices, and those linking factor prices to incomes and incomes to aggregate demand.

A few simple interactions among cities might also be postulated. Included here would be relationships between factor output and input levels and relationships between aggregate demands and incomes in different cities.

Admittedly, this model is quite crude in its present conceptual form. However, its principal merit is that it can be made both analytic and operational with a more detailed specification. Such a model permits us to examine some fundamental questions about the economic growth of an urban system.

Perhaps the most important of these questions concerns the allocation of growth among cities in an urban region or system. Why do some cities grow faster than others in income, in employment, or in output? Is there a tendency for large cities to gain an increasingly large share of the system's growth? Why do people migrate? Why does industry locate as it does? Does proximity to a fast-growing city help or retard the growth of neighbouring cities?

2. Sources of Growth and Factor Shifts

It is useful to measure economic growth in terms of changes in the physical output of goods and services. In an abstract sense, a neo-classical production function relates output, Q, to the physical inputs of capital, K, labour, N, and technology, A.

(1) $Q = F(K,N,A)$

If we define growth as the relative change in physical output, then it can be derived from (1) that

(2) $q = ak + bn + g$

where, using 'd' to denote a differential,

(3) $q = d \ln Q \quad k = d \ln K \quad n = d \ln N \quad g = d \ln A$

(4) $a = (\partial F/\partial K)/(Q/K) \qquad b = (\partial F/\partial N)/(Q/N)$

and assuming that 'g' is Hicks-neutral technical change.[3]

Since both 'a' and 'b' have been observed to be relatively constant, at least in the short run, the rate of growth of output, q, is linearly related by (2) to the rate of growth of capital, k, of labour, n, and of technical change, g.

There may be cases where 'g' might be expected to vary from city to city but, where the region lies totally within a single political jurisdiction with an open economy, it is often assumed that 'g' is everywhere the same. Thus, the relative rate of growth of output is determined by the rates of growth of capital and labour.

Labour growth, at full employment, can be attributed to either natural increase or net in-migration. Given a regionally steady rate of natural increase, the directions and magnitudes of migration flows become crucial to the determination of output growth rates. While much research effort has been expended on the determinants of migration flows, there appears to be some question as to whether migration induces or is induced by economic growth.

This migration-growth causality question is mirrored by questions concerning the causes of capital growth. If the capital growth rate, in

---

[3]Hicks-neutral technical change being a change in technology which does not affect the factor shares of either input.

other words industrial location, is influenced by the labour growth rate, this is equivalent to stating that net in-migration promotes economic growth. But, if capital growth affects the labour growth rate, this is equivalent to stating that net in-migration is influenced by economic growth. These two theories of economic growth may be termed the capital-shift and labour-shift hypotheses respectively.

There exist relatively few bases from which current models of urban systems are drawn. Most, both at the interurban and intraurban level, are based, in the main, on the labour-shift hypothesis. More recently, a capital-shift hypothesis has been formulated and some models have been designed on this basis. Also, a few models have been formulated which allow for a combination of these two hypotheses. Aside from this, there exist only scattered examples of other bases.

In the next few sections, much attention will be devoted to the labour-shift and capital-shift formulations. Currently-applied models will be described and their shortcomings noted. Also, some suggestions for improvements and future research will be suggested.

3. Labour-Shift Theory and Urban Growth

The export-base hypothesis is the most common formulation of a labour-shift hypothesis. Over the years, criticisms of this hypothesis in one form have repeatedly led to its demise only to be resurrected shortly thereafter in a revised version. What captures the recurring interest of researchers is the apparent theoretical simplicity of the hypothesis, quite often the same characteristic which deters those who try to use it empirically. However, a brief review of the various forms which this hypothesis has taken serves to point out that this theory, in

all its guises, suffers from a number of damaging deficiencies.

3.1. A Simple Export-Base Hypothesis

In one of its simpler forms, the export-base hypothesis recognizes two sectors of employment in a city at full employment, N, basic or export-oriented employment, B, and nonbasic or local oriented employment, S.[4] The kernel of this theory lies in the hypothesis that nonbasic employment, S, is linearly related to the aggregate population, P, of the city. Another hypothesis is that aggregate population, P, of the city is linearly related to total employment, N, of the city. In review, these conditions state

(5) $N = B + S$

(6) $S = \alpha_0 + \alpha_1 P \qquad 0 \leq \alpha_1 \leq 1$

(7) $P = \beta_0 + \beta_1 N \qquad \beta_1 \geq 1$

The definition (5) and the hypotheses (6) and (7) can be shown to assert

(8) $N = \gamma_0 + \gamma_1 B$ where

(9) $\gamma_0 = (\alpha_0 + \alpha_1\beta_0)/(1 - \alpha_1\beta_1) \qquad \gamma_1 = 1/(1 - \alpha_1\beta_1)$

Given that $\alpha_1\beta_1 < 1$, it is seen from (8) that total employment, N, is some multiple (greater than unity) of basic employment.[5]

Richardson (1969) identifies three main criticisms of this model. (1) It is difficult to differentiate empirically between basic and nonbasic employment. (2) The value of γ_1 is not a true constant since it changes

---

[4] Exports are those products shipped from the city to any other place.

[5] The condition that $\alpha_1\beta_1 < 1$ states that with a unit increase in population, the marginal change in service employment, α_1, must be less than or equal to the marginal change in total employment, $(1/\beta_1)$.

both with time and city size. (3) The model fails to analyze expenditure relationships, particularly those concerning savings and imports.

Of the three, the second criticism is most apparent in the model outlined. It is not difficult to see how β_1, in (7), might vary both with the age-sex composition and with the labour force participation rates of age-sex group in the population. It is not difficult to see either how α_1, in (6), might vary with changes in income, the demand for local-oriented goods and services, and the relative productivities of employees in basic and nonbasic industries. Thus, γ_1 cannot be expected to remain stable. Further, as has been noted by some observers, γ_1 is most stable for municipalities where α_1 is quite small. As can be seen from (9), the smaller α_1 is relative to β_1, the smaller the effects on γ_1 of proportional variations in either or both α_1 and β_1.

The following is concluded about the export-base hypothesis to this point. The exact mechanics of the model are difficult to see because of a number of complicated assumptions about prices, incomes, local and export demands, productivities, and demographic variables. In effect, the model assumes that there is a direct linkage between employment in basic and nonbasic sectors and that a sufficient amount of capital investment and labour can always be found to maintain this linkage.

If (8) is to be thought of as an approximation of reality, the question of how basic employment, B, is determined begs to be answered. If there is a clear division between basic and service employment, as this model presumes, then basic employment cannot be functionally related solely to the population of the urban centre. This argument holds for two reasons. First, if both employment sectors are related to population size, how could one functionally distinguish which of the two sectors any group of firms

belongs to. Secondly, if both basic and service employment are related solely to population size, (6) may very well be indeterminate.

This second reason is a critical one. Suppose that population is actually related to total employment, as hypothesized in (7) and suppose that both basic and nonbasic employment are linearly related to population size. Then, generally any population size at all can be supported because the total level of employment would adjust exactly to the requirements of that population.[6] Thus, indeterminacy arises.

---

[6] This result holds generally only where there exists at least one feasible population level. Since basic and service employment levels are linearly related to population size, we have

(a) $B = \alpha_0 + \alpha_1 P \qquad S = \beta_0 + \beta_1 P \qquad N = B + S$

and since population is linearly related to employment, we have

(b) $P = \gamma_0 + \gamma_1 N$

By substituting (a) into (b), it can be derived that

(c) $P = \gamma_0 + \gamma_1 (\alpha_0 + \beta_0) + \gamma_1 (\alpha_1 + \beta_1) P$

There exist three sets of solutions to this equation. First, there exists the case where

(d) $\gamma_1(\alpha_1 + \beta_1) \neq 1$

in which the population, P, is fixed at a constant value of

(e) $P = (\gamma_0 + \gamma_1(\alpha_0 + \beta_0))/(1 - \gamma_1(\alpha_1 + \beta_1))$

There exists the second case where

(f) $\gamma_1(\alpha_1 + \beta_1) = 1$ and $\gamma_0 + \gamma_1(\alpha_0 + \beta_0) \neq 0$

for which there is no feasible solution. Finally, there exists the case

(g) $\gamma_1(\alpha_1 + \beta_1) = 1$ and $\gamma_0 + \gamma_1(\alpha_0 + \beta_0) = 0$

for which any value of P represents a solution to (c). Given the derivation of (c), this last case seems to be the most appropriate.

Faced with the indeterminacy dilemma, there are only two recourses. Either one assumes that population growth is determined outside the export-base model or else one hypothesizes that basic employment, B, is determined by variables other than population size. Export-base theorists, as their name would imply, have chosen the latter recourse.

Following the line of attack of the export-base theorists, one must now ask why it is that output is produced in one city, say C_a, and shipped to another, say C_b. The most general response is that city C_a has some sort of locational advantages. However, faced with the indeterminacy dilemma, the export-base hypothesis demands that these locational advantages consist of more than just a potential labour pool. Capital investment and labour requirements in the export sector must not, to repeat, be solely reliant on population size.

Thus, it becomes evident that the export-base hypothesis is necessarily a labour-shift theory. An increase in the demand for exportable output leads to an increase in capital investment in the export-sector regionally. The amount of capital investment occurring at each urban centre cannot be such as to make the resulting export sector employment dependent on population size solely. Therefore, the capital growth rate in the export sector cannot be totally dependent on the population growth rate, the basis for a capital-shift theory.

However, an increase in export sector employment causes a multiple increase in total employment. If the natural increase in the local labour supply is not sufficient to meet these demands, migration occurs to satisfy the excess. There is no migration within the model other than that which must occur to fulfil these employment demands. There is no means by which new employment can be created from these migration flows since all such

flows are already committed to meet current employment demands. Thus, capital growth rates cannot be dependent on population size and cannot be induced by migration.

3.2. Czamanski's Models

In a series of papers, Czamanski proposed some extensions to the export base approach to overcome some of its shortcomings. An examination of these extensions provides a measure of the direction of research on the export-base theory.

Czamanski (1964) initially presented a slight extension to the model represented by equations (5) through (9). In this extension, he subdivides basic employment, B, into geographically-oriented industry employment, G, and complementary industry employment, C.

(10) $B = G + C$

By a geographically-oriented industry, Czamanski meant any sector whose choice of location is principally determined by non-population considerations Complementary industries include any industry which must locate near a geographically-oriented industry. It is hypothesized that there is a linear relationship between employment in the two industry sectors.

(11) $C = \varepsilon_0 + \varepsilon_1 G$

The solution to equations (5), (6), (7), (10), and (11) now is

(12) $N = \gamma'_0 + \gamma'_1 G$ where

(13) $\gamma'_0 = \gamma_0 + \gamma_1 \varepsilon_0 \qquad \gamma'_1 = \gamma_1(1+\varepsilon_1)$

On the surface, equation (12) appears to be no different from the original version (8). What Czamanski actually achieved by this extension was to place an emphasis on which kinds of industry should be basic and which nonbasic. Like Ullman and Dacey (1960) with their minimum

requirements technique, Czamanski was trying to overcome Richardson's first criticism, the delineation of industry sectors.

It is doubtful whether Czamanski's extension represents a significant improvement. Although he went to great lengths to distribute industries among his three sectors, it is clear that such distributions themselves are a function of technology and prices so that they cannot be expected to remain constant into the future, an important condition if the model is to be used for forecasting. Further, it is not entirely clear that Czamanski's delineation of sectors makes it any easier to identify which borderline industries belong where.

Czamanski (1965) proposed a second improvement to the export-base hypothesis in which he introduced lead-lag relationships between employment and population growth. The particular lengths of leads and lags are based on his work with forecasting the growth of the Baltimore SMSA and are not based upon more than qualitative theoretical arguments.

In this improvement, the population-employment relationship (7) becomes

$$(14)\quad P_t = \emptyset_0 + \emptyset_1 N_{t-2}$$

where the subscript refers to a year. Thus, (14) asserts the hypothesis that population growth follows employment growth with a lag of two years. Equations (5) and (10) are now represented by the current identity

$$(15)\quad N_t = G_t + C_t + S_t$$

From (6) is made the new hypothesis that local-oriented employment growth follows population growth with a lag of one year.

$$(16)\quad S_t = \alpha_0 + \alpha_1 P_{t-1}$$

Finally, equation (11) maintains its current form so that

(17) $C_t = \varepsilon_0 + \varepsilon_1 G_t$

The system of equations (14) through (17) reduces to the third-order difference equation

(18) $P_t = \Psi_0 + \Psi_1 G_{t-2} + \Psi_2 P_{t-3}$ where

(19) $\Psi_0 = \emptyset_0 + \emptyset_1(\alpha_0 + \varepsilon_0)$ $\Psi_1 = \emptyset_1(1 + \varepsilon_1)$ $\Psi_2 = \emptyset_1 \alpha_1$

Menchik (1971) terms (18) a dynamic equilibrium-seeking model which has an equilibrium solution associated with a particular level of G_t, say $G_t = G$, of

(20) $\lim_{e \to \infty} P_e = (\Psi_0/(1 - \Psi_2)) + (\Psi_1/(1 - \Psi_2))G$

for $0 < (1 - \Psi_2) < 2$

This equilibrium solution is equivalent in form to the solutions of the earlier versions of the economic base model, namely (8) and (12).

In effect, (18) is merely a partial-adjustment model in which population growth tends toward an export-base solution but in which the structural parameter ψ_2 and the deviation between the current economic base solution and the lagged population size serve to determine how quickly the population size tends toward that solution.

Casting the export-base hypothesis as a partial-adjustment model is a small but by no means trivial extension. Given that there is some kind of momentum or inertia to population growth, the hypothesis may be made more realistic by this extension.

However, this extension does not overcome any of the three main criticisms of Richardson - the difficulty in differentiating between basic and nonbasic sectors, the variability of the basic employment multiplier, and the lack of expenditure relationships. Also, this extension does not offer any possibility for a capital-shift theory.

Czamanski (1969) proposed a third revision to the export base hypothesis. In this final revision, he dealt with an annual growth model for a region, Nova Scotia, instead of an urban area but an urban model would be quite similar to this. The Nova Scotia model is quite complex, containing 31 structural and 23 definitional equations in 54 current endogenous variables.

However, by using a few simplifying assumptions, and by concentrating on just the employment and population growth variables, it is possible to arrive at a remarkably simple and telling representation of Czamanski's model. Let us assume (i) that the labour force is linearly related to population size and that there is a zero rate of natural increase.[7] Let us assume that (ii) capital stocks adjust perfectly to demand conditions.[8] Now, the part of Czamanski's model concerned with population and employment growth may be represented as

(21) $N_{1,t} = \alpha_{10} + \alpha_{11}FZ_t$

(22) $N_{4,t} = \alpha_{20} + \alpha_{21}UP_t$

(23) $N_{5,t} = \alpha_{30} + \alpha_{31}DFS_t + \alpha_{32}t$

(24) $N_{3,t} = \alpha_{40} + \alpha_{41}Y_{t-4} + \alpha_{42}SZ_t$

(25) $N_{2,t} = N^o_{2,t} \qquad N_{6,t} = N^o_{6,t}$

(26) $N_t = N_{1,t} + N_{2,t} + N_{3,t} + N_{4,t} + N_{5,t} + N_{6,t}$

---

[7] Czamanski used a cohort-survival method to forecast births and deaths by age and sex groups and applied a participation rate to the population in each group to derive the total labour force.

[8] Czamanski hypothesized a more-realistic partial adjustment equation but our simpler formulation makes the model easier to interpret.

$$(27) \quad P_t - P_{t-1} = \alpha_{50} + \alpha_{51}(N_{t-1} - \alpha_{52}L_{t-1})$$

$$(28) \quad L_t = \alpha_{60} + \alpha_{61}P_t$$

where the numeric subscript of N refers to employment in (1) agriculture, forestry, and fishing, (2) mining, (3) manufacturing, (4) commercial services, (5) government including military personnel, or (6) iron and steel, and where

	FZ	is the ratio of commercial to all farms,
	UP	is urban population,
	DFS	is payments to military personnel,
	t	is time (in years),
	Y	is an economic activity index measure,
	SZ	is an index of size of plant,
	P	is population,
and	L	is labour force.

Equations (21) through (23) correspond to Czamanski's (S.4) through (S.6) respectively. Equation (24) is adapted from Czamanski's equations (S.7), (S.9), (S.10), and (D.2) using assumption (ii) above. Note that all variables on the right-hand-side of equations (21) through (24) are purely exogenous so that there are no feedbacks within the model from activity to employment other than those directly specified here.[9]

Equation (25) merely specifies that $N_{2,t}$ and $N_{6,t}$ are exogenously given. Equation (26) states the employment balance and corresponds to Czamanski's (B.1). Equations (27) and (28) are adapted from Czamanski's equations (D.11) through (D.15) and (S.29) and using assumption (i) above.

---

[9] This includes the variable Y_{t-4} which is a lagged exogenous variable within the model.

By examining equations (21), (23), (24), and (25), it becomes clear that the non-service employment level within the model is determined by variables which are exogenous to the model. While exact specification of these exogenous relationships make the model more sophisticated, this is nonetheless merely a re-assertion of the export-base hypothesis; non-service, or basic, employment is exogenously given and service, or local, employment is functionally related to population size.

This conclusion is especially surprising in view of the fact that one major reason for constructing econometric urban growth models is to overcome Richardson's third criticism of economic base theory, its inability to handle imports, savings, or other expenditure leakages. Astoundingly, because of the exogenous determination of nonservice employment, Czamanski's model does not come on inch closer to overcoming this criticism than did the simplest economic base model outlined earlier!

About all that this model really contributes to the export-base hypothesis is summarized in (27) and (28). Equation (27) hypothesizes that net in-migration (or out-migration) is something like a partial-adjustment process where a level of population is gradually achieved which is consistent with the amount of labour required in (26). Thus, (27) and (28) in this Nova Scotia model are operationally very similar to the mechanism, described by (18), contained in Czamanski's Baltimore model. In fact, it is not clear that the Nova Scotia model contributes anything more to the export-base hypothesis.

3.3. Glickman's Philadelphia Model

Czamanski's models have provided a useful means of summarizing the development of export-base models during the period of 1964 through 1969. Glickman's 1971 model of Philadelphia represents one of the more recent

and most sophisticated attempts to model the aggregate growth of a single urban area.

Glickman (1971) proposed a model consisting of 17 structural and 9 definitional equations, somewhat smaller than Czamanski's Nova Scotia model. Unlike Czamanski however, Glickman specifically introduces factor and product prices into export-base theory. Again, this appears to be an attempt to overcome Richardson's third criticism, the lack of expenditure relationships.

Glickman's model can be summarized into an abbreviated set of nine structural and definitional equations as follows.

$$GRP_t = \alpha_{10} + \alpha_{11} GNP_t + \alpha_{12} PY_t \tag{29}$$

$$N_t = \alpha_{20} + \alpha_{21} GNP_t + \alpha_{22} PY_t + \alpha_{23} t \tag{30}$$

$$WM_t = \alpha_{30} + \alpha_{31}(L_t - N_t) + \alpha_{32} WMUS_t \tag{31}$$

$$PR_t = \alpha_{40} + \alpha_{41}(WM.N/GRP)_t + \alpha_{42} N_t \tag{32}$$

$$WR_t = WM_t / PR_t \tag{33}$$

$$PY_t = WR_t N_t + NWY_t \tag{34}$$

$$NWY_t = \alpha_{50} + \alpha_{51} GRP_t \tag{35}$$

$$L_t = \alpha_{60} + \alpha_{61} N_t + \alpha_{62} t \tag{36}$$

$$P_t = \alpha_{70} + \alpha_{71} L_t + \alpha_{72} t \tag{37}$$

where

GRP is gross regional product[10]

GNP is gross national product[10]

PY is regional personal income[10]

N is total regional employment

---

[10] in constant dollars.

	WM	is the regional money wage rate[11]
	L	is the regional labour force
	WMUS	is the national money wage rate[11]
	PR	is the regional consumer price index
	WR	is the regional real wage rate[10]
	NWY	is regional non-wage income[10]
	t	is time in years
and	P	is regional population

and where all variables are measured within the Philadelphia region other than those directly specified as national.

In (29) and (30), gross regional product and regional employment are related to both gross national product and to regional personal income, reflecting the basic and nonbasic sectors respectively. The regional money wage rate is linked to the national money wage rate and to the regional labour shortage or surplus in (31). In (32), the regional price level is related to the employment level and to average unit labour costs. Two definitions are formulated in (33) and (34). In (35), regional nonwage income is related to gross regional product. Regional labour force is a function of regional employment and time in (36) and in (37) regional population is related to regional labour force and time.[12]

---

[11]in current dollars.

[12]Referring to Glickman (1971; pp19-20), the above equations are derived as follows:

Equation	from Glickman's
(29)	(1), (2), (3), (4)
(30)	(1), (2), (3), (4), (7), (8), (9), (10)
(31)	(11), (12), (13)
(32)	(14)
(33)	(15), (16), (17)
(34)	(19)
(35)	(18)
(36)	(20)
(37	(21)

Let us now initially assume that prices and wages are fixed. From (29), (30), (34), and (35), it can be derived that regional personal income is linearly related to gross national product such that

(38) $PY_t = \beta_{10} + \beta_{11}GNP_t$ where

(39) $\beta_{10} = (WR_t\alpha_{20} + \alpha_{50} + \alpha_{51}\alpha_{10})/\gamma_0$ $\quad\beta_{11} = (WR_t\alpha_{21} + \alpha_{51}\alpha_{11})/\gamma_0$

$\gamma_0 = 1 - WR_t\alpha_{22} - \alpha_{51}\alpha_{12}$

Now, using (30) and (38), it can be deduced that regional employment is linearly related to gross national product such that

(40) $N_t = \beta_{20} + \beta_{21}GNP_t + \alpha_{23}t$ where

(41) $\beta_{20} = \alpha_{20} + \alpha_{22}\beta_{10}$ $\quad\beta_{21} = \alpha_{21} + \alpha_{22}\beta_{11}$

The employment determination equation, (40), together with the labour force and population equations, (36) and (37), provide a simplification of Glickman's model under the assumption of constant wages and prices.

(40) $N_t = \beta_{20} + \beta_{21}GNP_t + \alpha_{23}t$

(36) $L_t = \alpha_{60} + \alpha_{61}N_t + \alpha_{62}t$

(37) $P_t = \alpha_{70} + \alpha_{71}L_t + \alpha_{72}t$

In this simplification, gross national product, through its effect on the basic sector, determines regional employment which determines regional labour force which determines regional population. This is nothing more than the seemingly ever-present export-base hypothesis.

Suppose that the assumption of constant wages and prices is dropped so that β_{20} and β_{21}, in the above system, are variable. Assume a random shock, say an unexpected increase in in-migration, in equation (36) which increases the labour force, L. Through (37), there is the initial obvious consequence that population, P, is also increased. However, the money

wage rate, WM, falls since α_{31} in (31) is expected to be negative. The consumer price index, PR, will now decrease since α_{41} is expected to be negative also. Since one might well expect the money wage rate to decrease proportionately more than the consumer price index, the latter being more tightly tied to national trends, it is anticipated that the regional real wage, in (33), and regional personal income, PY, would decline. As a consequence of this, gross regional product, GRP, would decline in (29) as would regional employment, N, in (30). Thus, β_{20} nd β_{21} respond, in the simplified system above, to permit labour force change to have an effect on employment.

This last illustration also serves to point out something incorrect about the mechanics of this model. A <u>ceteris paribus</u> increase in labour force has been shown to lead to an initial (if not final) effect of a decline in total employment. Certainly, one might well have expected an exactly-opposite effect.

If one re-examines the model carefully, there would appear to be a mis-specification of the labour market equations (30) and (31). In (30) regional employment is given as a function of GNP and PY which are measures of national and regional aggregate demand. There is, however, no relationship between regional employment and the regional wage itself. Thus, there is no means by which a reduction in the wage rate can lead directly to an increase in total employment.

This mis-specification makes it difficult to evaluate the contribution of this model to the export-base hypothesis. If, however, one amends (30) to the form of

$$(42)\quad N_t = \alpha_{20} + \alpha_{21}GNP_t + \alpha_{22}PY_t + \alpha_{23}t + \alpha_{24}WR_t$$

as an example, it is evident that Glickman has circumvented Richardson's

third criticism, the lack of expenditure analysis. Now, factor and product prices as well as outputs (and therefore expenditures) have been directly taken into account.

Now, however, it is also clear that the substitution of (42) for (30) results in a fundamentally different model for another reason. Previously, there could be only one level of total employment and population associated with a particular level of exports and export employment. In this amended form of Glickman's model, it is now possible to initially have a surplus population, in the sense that labour force exceeds employment at a given wage, and yet to have wages adjust so that labour force and employment gradually become equal.

This being the case, Glickman's amended model evidently represents a major step away from the export-base models described earlier. Aggregate employment growth can now be affected by the size of population so that, to some degree, a capital shift hypothesis has been introduced. However, it has to be appreciated that, within the model, population is still strongly tied to employment levels. Population growth affects employment growth only in random shocks: there is still no systematic trend to population growth other than as a multiplicative response to export growth.

4. Growth in an Urban System and Labour-Shift Theory

To this point, we have summarized some of the developments in export-base theory. This has by no means been an exhaustive survey of literature on the hypothesis but has served to outline how, and in what directions, research has been headed. Thus far, the models examined have been concerned with the growth of a single urban centre. The concept of an urban system has yet to be introduced as an aspect of a growth model.

There appears to be a considerably smaller literature on economic growth within a system of urban centres than on the growth of single urban areas. In part, this may be attributed to the nice property of single-centre models that they tend to be relatively open models.[13] Urban systems models, however, tend to be somewhat less open although the degree to which they are closed depends on the magnitude of the system.

A troublesome feature of a relatively closed economic growth model is that all factor movements must be accounted for twice; once as an outflow from a 'losing' centre and once as an inflow to a 'gaining' centre. The result is that generally factor allocation mechanisms have to be nonlinear because of this counting constraint and this poses significant analytic and conceptual problems. In an effort to avoid these problems, researchers have generally relied on models which are very rigid either in terms of their assumptions or in their treatment of factor balances.

One approach here is to use an interurban input-output model, equivalent to the well-known interregional input-output model.[14] Such a model is very rigid, however, in terms of the number of assumptions which have to be made regarding production, consumption, and trade flows.

It is clear, moreover, that an interurban input-output model with an endogenous household sector is nothing more than a detailed export-base model extended to cover an urban system. Thus, they face the same criticisms

---

[13]An open model is one in which external relationships play an important role. A closed system is typified by extensive balance equations and a double-accounting principle. Double accounting in an open model is very difficult because of the external relationships. Open and closed models, of course, represent extremes as any particular model lies somewhere between the two.

[14]Refer to Miernyk (1965) for a survey of this model.

as other versions. The emphasis is slightly different here though since assumptions about basic versus nonbasic industries, and about expenditure relationships, are made quite explicitly in terms of coefficient values. Unlike earlier export-base models where the same assumptions are 'hidden' in generalized relationships, the explicit assumptions in this model can be varied in a more meaningful sensitivity analysis.

There are, however, two additional criticisms, beyond Richardson's three, which apply to the interurban input-output approach. The first is the problem of dimensionality, of handling large or even moderate numbers of cities, given that the number of coefficients increases with the square of the number of cities. This point is well-noted in the literature and need not be emphasized further.

The second additional criticism of this model is that the coefficients are too disaggregated to remain stable over a very long period of time. This is felt to be especially the case with the trade coefficients. Wickeren and Smit (1971) have been considering this specific problem in relation to their proposed interregional input-output model of the Netherlands. They propose to use a 'dynamic attraction model' to determine the trade coefficients separately and then to use the input-output model with these coefficients to determine production, consumption, and trade patterns.

Regardless of whatever contributions may eventually follow from the work of Wickeren and Smit and others, it would appear that the interurban input-output model is too demanding in terms of data, too rigid in terms of assumptions, and too lacking in population-wage-employment feedback mechanisms to be useful as an interurban growth model.

5. Capital Shift Approach and Urban Growth

To this point, we have examined labour-shift approaches to urban economic growth where it is assumed that economic growth in the form of employment determines the population which can be supported at an urban centre. An alternative approach, as outlined earlier, is via a capital-shift theory in which it is hypothesized that population growth is itself the primary cause of economic and employment growth. In this regard, Borts and Stein (1964) undertook a large study of regional economic growth in the United States and were among the first to note that employment growth might be primarily determined by population growth.

Since the work of Borts and Stein, a number of capital-shift models have been formulated. These have typically been mixed models allowing for both capital and labour shifts. A selected set of these models are analyzed in the following sections.

5.1. Niedercorn's Model

Niedercorn (1963) presents a two-part dynamic model of an urban centre in which a first part is concerned with aggregate urban growth and the second with the city-suburb division of that activity. Here, we will be concerned with the first part.

In that part of the model, Niedercorn recognizes the two familiar forms of employment; basic (B) and nonbasic (S). It is hypothesized that there is a desired level of basic employment, B_t^*, which is proportional to the current population of the urban centre, P_t.

$$(43) \quad B_t^* = \alpha_{11} P_t$$

This represents a much different hypothesis than is found in export-base theory where basic employment is related to export levels.

Niedercorn then hypothesizes that the actual level of basic employment, B_t, is given by a partial adjustment model

(44) $R_t - R_{t-1} = (1 - \alpha_{21})(\alpha_{22}R^*_{t-1} - R_{t-1})$ where

(45) $R_t = (B_t - B_{t-1})/B_{t-1}$ $R^*_t = (B^*_t - B_t)/B_t$

The final hypothesis of the first part of this model states that the population growth rate, S_t, is linearly related to the basic sector employment growth rate, R_t.

(46) $S_t = \alpha_{30} + \alpha_{31}R_t$ where, by definition,

(47) $S_t = (P_t - P_{t-1})/P_{t-1}$

It has proven difficult to analyze the properties of the model contained in (43), (44), and (46). In (44), it is asserted that basic employment growth rates tend toward an equilibrium. However, that equilibrium relates to a previous time period and in (46) it is seen that the current equilibrium is always changing.[15]

In fact, given that α_{30} and α_{31} are positive, it would seem that there is no steady state solution to the system. The faster basic employment increases to catch up with its lagged desired growth rate in (44), the faster the current population growth rate increases in (46), and the faster the desired basic employment growth rises in (43). In this sense, the system is unstable.

---

[15] It was noted in section 3.1. and expanded in footnote 6 that basic sector employment in the export-base model could not be linearly related to population size. There, the consequence of relating the two variables, in current form, was to permit the export-base model to have an infinite number of solutions. A single solution is permitted in this model because of the time-lag structure of the model. However, a time path of single-year solutions tends to infinity for $\alpha_{30} > 0$.

The fact that Nidercorn's model has feedback relationships between population and basic employment growth is enough to justify calling it a capital-shift model. However, it is not clear that this model is any more theoretically-satisfying than were the earlier export-base models. Still applicable are Richardson's first and third criticisms of the export-base model; the difficulty of distinguishing between basic and nonbasic sectors, and the lack of expenditure analysis.[16]

In addition, this model as it stands does not seem to be any more amenable to development as an interurban model than did the export-base models. Clearly, at least equation (46) has to be respecified to allow for differential growth rates for urban centres attributable either to interactions such as migration and commodity flows or to other nonemployment considerations. There is some promise however that this model can be adapted as an interurban model.

Suppose now that, for urban centre 'i', (46) is replaced by

$$(48) \quad S_{i,t} = \alpha_{30} + \alpha_{31}R_t + \alpha_{32}A_{i,t-1} \quad \text{where}$$

$$(49) \quad A_{i,t} = \sum_{j=1}^{N} f_{ij}P_{j,t}$$

In these equations, f_{ij} is the functional distance between centres 'i' and 'j' and $A_{i,t}$ is a measure of the accessibility of centre 'i' at time 't'.

We would hypothesize that α_{32} is positive in (48). In part, α_{32} is positive because accessibility is widely regarded to be a benefit and as such a ceteris paribus increase in accessibility makes a place more desirable to live in and therefore helps to increase its growth rate.

---

[16] Niedercorn attempts to get around the first by insisting that the basic sector is strictly composed of manufacturing.

Secondly, accepting that migration flows are affected by distance, the accessibility measure serves to indicate the ease with which in-migration and, therefore, population growth can be encourages at centre 'i'. Since the accessibility measure is related to both populations within the region and their spatial distribution in terms of the networks linking them, we now have the crude semblance of a regional mechanism operating in (48).

Suppose now that an urban system is characterized by (43), (44), and (48). Suppose also that, for one centre 'i', equation (48) is subject to a series of random shocks which cause centre 'i' to grow rapidly in the space of a few years. According to (48), this will cause an increase in the growth rates of nearby centres because of their accessibility to 'i'. This, in turn, will encourage further growth at 'i' because its own accessibility is increasing as the populations of nearby centres increase. Just how significant these interurban impacts are numerically can only be determined when α_{32} and the associated elasticities are calculated.

Finally, one can note the significance of α_{31} in Niedercorn's model. If α_{31} is near zero, using either (46) or (48), population growth becomes independent of employment growth. In this case, the model becomes a pure capital-shift theory. A value for α_{31} which is greater than zero implies a mixed capital-shift and labour-shift model.

5.2. Muth's Model

Muth (1968, 1971) developed a model of differential growth in large cities based on the regional theories of Borts and Stein. In it, he assumes that a city produces two commodities; an export good, X, and a domestic good, Z. Both goods are produced, with constant returns to scale, using the two factors, labour (N) and capital (K). This implies that

$$(50)\quad X^* = \rho_{nx} N_x^* + \rho_{kx} K_x^* \qquad \rho_{nx} + \rho_{kx} = 1$$

$$(51)\quad Z^* = \rho_{nz} N_z^* + \rho_{kz} K_z^* \qquad \rho_{nz} + \rho_{kz} = 1$$

where

ρ_{ab} is the elasticity of B with respect to A

N_a is employment in sector A

K_a is capital stock in sector A

and $*$ denotes a logarithmic differential (e.g., $X^* = d\ \ln X$)

Muth earlier derived the demand function for the factor inputs, N and K. Assuming that wages are the same in both sectors and that the price of capital is given exogenously, the conditions for the export sector only are the following.

$$(52)\quad \rho_{kx}(N_x/K_x)^* = \sigma_x(p_x/w)^*$$

$$(53)\quad \rho_{nx}(K_x/N_x)^* = \sigma_x p_x^*$$

Where σ_x is the elasticity of factor substitution in industry X,

p_x is the price of good x,

and w is the universal wage rate.[18]

---

[18] The derivation of (52) and (53) is fairly lengthy and only (52) will be developed as an example.

Let (a) $X_n = \partial X/\partial N \qquad X_{nn} = \partial^2 X/\partial N^2 \qquad X_{kn} = \partial^2 N/\partial K\partial N$

Then, (b) $d\ X_n = X_{nn} dN + X_{kn} dK$ by differentiating (a)

and (c) $d\ln X_n = (NX_{nn}/X_n) d\ln N + (KX_{kn}/X_n) d\ln K$ from (b).

But, (d) $X = NX_n + KX_k$ by Euler's equation,

so (e) $X_n = X_n + NX_{nn} + KX_{nk}$ from differentiating (d)

and hence (f) $X_{nn} = -(K/N)_x X_{nk}$ from (e).

From (52) and (53) can be derived the linkage between the wage rate and the export good's price

$$(54) \quad w^* = (1/\rho_{nx})p_x^*$$

Since there are constant returns to scale and an infinitely elastic supply of capital, the wage rate is independent of the scale of production. In effect, employment at a given wage rate is limited only by the supply of labour at that price.

Muth then hypothesized that employment, N, is related to the level of migration, M, to the wage rate, W, and to the natural increase in population, n.

$$(55) \quad N^* = \alpha_{11}M + \alpha_{12}w^* + \alpha_{13}n$$

Finally, migration is linked to the level of employment, the wage rate, and the rate of natural increase.

$$(56) \quad M = \alpha_{21}N^* + \alpha_{22}w^* + \alpha_{23}n$$

Schematically, Muth uses the causal structure outlined in Figure 1. Export prices are seen to determine wages in the export sector. Wages in the two sectors are assumed to move toward parity. It is the aggregate wage level which determines the level of employment and the level of migration in

---

Now (g) $d\ln X_n = -(KX_{nk}/X_n)d\ln N + (KX_{kn}/X_n)d\ln K$ from (c) and (f)

but (h) $\sigma_x = (X_n X_k)/(XX_{kn})$ from a well-known theorem,

so (i) $d\ln X_n = (\rho_{kx}/\sigma_x)(K/N)_x^*$ from (g) and (h).

If one assumes a competitive labour market, then the marginal productivity of labour is equal to its real wage or

(j) $X_n = w/p_x$

so (k) $(w/p_x)^* = (\rho_{kx}/\sigma_x)(K/N)_n^*$ from (i) and (j)

or (l) $\rho_{kx}(N_x/K_x)^* = \sigma_x(p_x/w)^*$

conjunction with the rate of natural increase.

Muth's model is much different from any others examined to this point. In the export-base model, aggregate employment is linked to export-sector employment. In this model however, aggregate employment is linked to export-sector prices. Further, in this model there is no limit to the amount of employment which can be generated at a centre, assuming that the city size is small enough not to affect export prices nationally.

Figure 1:

There are, however, a few basic problems associated with Muth's model. First, there is the old problem of the export-domestic sectoral distinction seems to have haunted most of the models to this point. Secondly, there is the problem that, like Niedercorn's model, this model does not allow for the concept of an urban system.

The latter problem might be overcome in the same way as the Niedercorn model. Equation (56) could be re-specified to include an accessibility variable so that

$$(57) \quad M_{i,t} = 21N^*_{i,t} + 22w^*_{i,t} + 23n_{i,t} + 24A^*_{i,t-1}$$

using subscripts to denote centre 'i' and time 't', and $A_{i,t}$ as defined in (49) with

$$(58) \quad P_{i,t} = P_{i,t-1} + n_{i,t} + M_{i,t}$$

The consequences of including (57) in the model are equivalent to those outlined for Niedercorn's model.

Finally, there remains one new criticism which is unique to Muth's model. By assuming that export prices are given nationally, Muth is denying some of the most basic concepts in spatial price theory. In his analysis, which was concerned with large cities, this assumption may not have posed significant problems. In a regional system of urban centres, however, export prices will vary depending on the characteristics of the regional markets in which each centre competes. It appears to be useful to think of export prices as being related both to national average prices and to the characteristics of the local regional markets in terms of competing suppliers and market demand. It is clear then that (54) needs to be re-specified although exactly how is not yet clear.

6. Conclusion

At the outset of this paper, it was argued that there are two sources of relative economic growth for an urban centre in an open region; labour force growth through natural increase or migration, and capital investment. What becomes essential to urban growth theory is then the reason why capital is invested and why people migrate. It quickly becomes clear that the two are really different ends of the same horse and the major problem is in determining which way the horse is moving. If capital investment is the leading end and migration follows, then a labour-shift theory is relevant. If migration leads, then a capital-shift theory is the appropriate one.

Further, it has been noted that many of the current economic growth models represent labour-shift theories exclusively. These include the

export-base theory in all its guises, various interregional and interurban input-output models, and several of the more complicated econometric models.

Of the models considered here, Glickman's model and the abbreviated interurban input-output model seem to offer the most hope for application to an urban system. Both, however, will require further revisions to overcome the criticisms noted.

While the notion of a labour-shift theory may have some attractiveness, it is not at all clear that a capital-shift theory is any less tenable. Two versions of models which permit both capital and labour shift hypotheses have been presented. Of the two, Muth's model is superior in terms of theoretical elegance although several criticisms of the model have been noted. The other model, by Niedercorn, although considerably less elegant, offers some promise with revisions because of its simplicity.

It appears at this point that we are quickly reaching the stage when urban systems can be modelled in a simple yet elegant way. Thus, we may shortly be able to begin answering the kinds of questions outlined in the introduction.

APPENDIX B

Part III: Urban Simulation Models

Eric Sheppard

Batty, M. 1971. Modelling Cities as Dynamic Systems. Nature 231: 425-428.

A model which attempts to integrate spatial interaction with the processes of change.

The Model

(a) Basic employment is an input in the model and generates service industry and population in decreasing amounts over time: "the economic base hypothesis."

Thus $\lim E = B(1+\lambda+\lambda^2 = \ldots\ldots+\lambda^n)$ $\quad 0>\lambda>1$

E = activity generated by input B

Also included are locational changes of basic employment within the city.

(b) Other activities are allocated as land use changes within the area. The model used here is the gravity formulation where the amount of activity depends on the locational attraction of the area. In this way, service employment and population are allocated.

Thus for population change:

$$\Delta P_j^{(t+1)} = \gamma \sum_i \Delta T_{ij}^{(t+1)}$$

$\Delta P_j^{(t+1)}$ = population change in j in time (t, t+1)

$\Delta T_{ij}^{(t+1)}$ = trip generation change from i to j in (t, t+1)

Where

$$\Delta T_{ij}^{(t+1)} = A_i^{(t+1)} \cdot \Delta E_i^{(t+1)} \cdot \delta_j^{(t)} \cdot (\sigma . L_j^{(t)} + (1-\sigma).F_j^{(t)})\ e^{-BC_{ij}^{(t)}}$$

and

$$A_i^{(t+1)} = [\sum_j \delta_j^{(t)} . (\sigma . L_j^{(t)} + (1-\sigma).\ F_j^{(t)})\ e^{-BC_{ij}^{(t)}}]^{-1}$$

$\Delta E_i^{(t+1)}$ = change in employment in i in time (t, t+1).

$\delta_i^{(t)}$ = Kronicker Delta (=0 when total capacity of a cell is exceeded).

$L_j^{(t)}$ = available land in j at time t.

$F_j^{(t)}$ = floorspace in j at time t.

$C_{ij}^{(t)}$ = cost of travel from i to j at time t.

γ, B, σ are constants.

(c) Basic employment that is dependent on location of present facilities in the area for its own location is estimated from:

$$\Delta Y_i^{(t+1)} = \sum_j \alpha_j X_{ij}^{(t)} + \sum_k \Theta_k \Delta Z_{ik}^{(t)}$$

$\Delta Y_i^{(t+1)}$ = change of basic employment in time (t, t+1).

$X_{ij}^{(t)}$ = independent variables (eg. total employment).

$\Delta Z_{ik}^{(t)}$ = changes in the k^{th} independent variable in time (t-1, t).

(d) An accounting procedure is used at the end of each generation to check that constraints are not being violated.

In this model it appears to take ten generations before the effect of basic employment on the growth of other sectors decreases to zero. So in any generation, the growth in all cells depends to varying degrees on the basic employment allocated in all the previous ten generations. So the model is always in disequilibrium at any point in time.

Validation of the Model

Based on the Reading area, England. The spatial structures explicit in the model allow for outputs of features like mean work trip length and service trip length as well as the basic variables.

Evaluation of the model under alternative policies is possible providing a criterion is available to compare the relative merits of alternative outcomes.

Aggregation

Spatial: areas (based on the census) of 3-12 sq. miles.
Temporal: arbitrary.
Social: the only split is of employment into basic and non-basic categories.

Comments

The model is at a very broad scale producing consequently coarse parameters as outputs, but at certain levels of aggregation this is no disadvantage.

Necessary data inputs are relatively low as the model relies on theoretical structure.

Only human activities are explicitly dealt with. Much of the structure of the city such as transportation routes and the physical age and quality of housing are not discussed.

* * *

Blumberg, D.F. 1971. The City as a System. Simulation 17: 155-167.

An attempt to look at the city as a whole, combining several models developed elsewhere to produce a model of the community.

Model

Three models are used in this development: The Penn-Jersey transportation model, the Pittsburgh land use allocation model, and the San Francisco housing renewal model. These give rise to the community model which is in four steps:

(1) A small area model describing each neighbourhood, the way of life, and secondary employment generated
(2) An area sub-model describing in more general terms the community as a whole
(3) A model of changes in individuals and groups: aging, deaths, education, etc.
(4) A household model describing household attributes: location, income, etc.

Each model's output is input to the next one; and the fourth model's output provides the input to the 1st model for the next generation. The model works for eareas of a size of less than 1 square mile.

It is also hoped to develop a model of public policy which will be integrated into the community model in order to evaluate the effects of alternative policies.

Results

A printout is demonstrated for a projection of Toronto in 1974 by sub-areas, but there is no attempt to check the pattern or in anyway test the model against the real world.

An example is also given, for a hypothetical city, of the effect of alternative schooling policies.

Comment

Attempts to create models of the city as a whole are extremely useful but they depend on the assumptions underlying the sub-models used. In particular it has been shown that the San Francisco model alone has shortcomings. It is difficult to say whether it would work better in combination with other models, especially as this paper does not give any explanation of the structure of the model.

Centre for Real Estate and Utban Economics. 1968. Jobs, People and Land. Special Report No. 6. University of California, Berkeley, California.

The Bay Area Simulation Study model; a modification of the Lowry model designed to test the effect of locating large industrial plants, on the allocation of commercial and residential activities.

The model projected employment, households and population; and disaggregated parameters, which Lowry used as aggregate measures of the entire system, by census tract. These were evaluated by base year relationships rather than as system-wide averages. As a result of initial trials this original model was modified and published in this paper.

* * *

Crecine, J.P. 1964. A Time Oriented Metropolitan Model for Spatial Location. Department of City Planning, Pittsburgh, Pennsylvania.

The Time Oriented Metropolitan model is a modification of the Lowry model, where rather than allocating all the city's activities, only change and growth in activities was allocated. Also households were disaggregated by income, housing characteristics and social characteristics. To implement this, activities were partitioned to stable and mobile activities - an exogenously determined partition. Also land uses are similarly partitioned into those which can and those which cannot change. This has not been operationalized.

* * *

Crecine, J.P. 1968. A Dynamic Model of Urban Structure. Santa Monica, California: Rand Corporation.

The TOMM model is modified such that variables are more disaggregated, the concept of site amenities is introduced, and zoning constraints are explicit. Also explicit recognition is made of the inertia in the system due to durability of physical structures. These changes are backed by careful theorising and explicit attention to computer applications.

Future attention is to be given to trying to get data for the model, as it is considered to be taken as far as is possible under the present structure.

* * *

Colenutt, R.J. 1969. Linear Diffusion in an Urban Setting: An Example. Geographical Analysis 1:106-114.

A model to simulate the growth of billboards along highways outside a city.

Model

The probability of a stretch of road getting a billboard is proportional to the amount of traffic.

Constraints

(a) A limit of 6 on the number of billboards per quarter mile.
(b) A removal constraint of billboards on the edge of the city as development occurs on these sites.
(c) A certain number of "contacts" have to be made with a landowner before he permits billboards on his land (determined by an arbitrary constant).

Evaluation of Model

Spectral analysis of residuals of an application to State College, Pennsylvania was used as a test of goodness of fit, and a cycle of large residuals on a wave length of 2.5 miles was uncovered; possibly related to other urban developments which occur at multiples of this distance from State College.

Comments

An extremely simple model of an empirical regularity. Such things as resistance of landowners to billboards are difficult to model. The model is based on a probability surface that is constant over time, which seems unrealistic as traffic flow is linked to urban development.

* * *

Cowan, P. et al. 1967. Approaches to Urban Model Building. Regional Studies 1:2:163-172.

A model to simulate the development of offices in a city.

Components of Model

(1) New offices = New Offices built + buildings converted to offices - offices demolished.

(2) Total number of new offices built as weighted average of the number built in the two previous time periods.

(3) Total number of buildings converted to offices is a similar average.

(4) Allocation of offices to each cell is according to an index of attractiveness: The number of offices in the cell divided by the total number in the city. The more attractive the cell the greater is the probability of offices going there. Actual allocation is decided by drawing random numbers.

(5) The probability of getting permission to build is determined, and random numbers are used to determine whether any particular office gets this permission.

(6) Random numbers are also drawn to determine the time to build the office (only when it is completed does it increase the attractiveness of the cell it is in).

Aggregation Levels

(1) Spatial: cells of 500 metres square.

(2) Temporal: one generation is one year.

(3) Social: all offices are of the same type.

Exogenous Variables

Empirical data on the probability of getting permission and of building an office in a given time period, the number of offices at the two previous time periods, and the proportion of offices demoslished are all needed.

Ability to Forecast

By iteration can test alternate land use and planning permission legislation.

Testing Results

Can be done assuming stochastic relations are constant, but not yet carried out.

Comments

No real spatial relations are included; for instance rate of demolition is the same in all cells. Offices are assumed to be of one type. The attractiveness index is only a simple proportional measure with no theoretical background. Supply and demand relations for the amount of office building are not included within the model. In general it is a very empirical model, based on extrapolation for many of its parameters.

* * *

Crecine, J.P. 1968. A Dynamic Model of Urban Structure. Rand Corporation: P3803 62pp.

A model of city land-use change based on the Lowry model.

Model

The model differs from the Lowry model in that instead of starting with a totally empty land area and developing a city on it given the location of non-population oriented industry, Crecine develops a model of incremented change through time. This is done by defining a certain proportion of industry and housing as mobile (so that each area can lose no more than this proportion of each activity in one time period), and relocating this as well as is possible with respect to the location of new exogenous industry. Constraints to location are based on zoning laws, minimum employment defined as necessary for setting up a business, and the amount of housing that can be built in one time period.

Households are divided up into different types, each of which has a preference for locating on a site based on work and shopping trip costs, rent, and site amenities (schools, etc.). The rent of any site depends on the general accessibility of that site to the rest of the urban area as a whole, and on the demand for housing there, relative to the supply in the previous time period. This would differ from the location cost of the site to households, as households consider only those work places in which that group is employed.

For population oriented industry, locational preference depends on accessibility to the market i.e. to the households.

The model proceeds in the following form:

(a) Feed in location of projected new exogenous firms by cell.
(b) Locate households as well as is possible with respect to this. The number of households depends on the new firms' employment.
(c) Locate new endogenous industry with respect to the market of households.
(d) Calculate and locate households necessary to provide the endogenous employment.
(e) Calculate and locate endogenous firms necessary to supply the new households.

Recycling occurs between (d) and (e) in even smaller iterations until the two elements are in equilibrium. Because of capacity constraints on cells, and because only a certain number of the urban activities can move in any one time period, the model is not likely to achieve the ideal locational equilibrium between firms and households. However the best possible compromise, given the circumstances is obtained.

Aggregation

Spatial:	By census tracts
Temporal:	Arbitrary
Social:	As many household and firm types as is practical given the data.

Testing and Evaluation

In order to use the model, parameters must be calculated for the distance decay function, for the functions of demand and rent elasticities with respect to site characteristics (for each household type), and rent elasticities with respect to demand relative to supply. Also needed is information on labour participation rates, employment needs for firms, and size of market served by the various endogenous firms. These data needs are large and as a result the model is still in its theoretical form. Crecine feels that theoretically the model can be taken no further until it is tested. It is hoped to apply the model, and linear functions are assumed so that calibration by regression is relatively easy.

Comments

The model assumes that a quasi-equilibrium is achieved within each generation, before any more exogenous employment comes in. The usefulness of this concept depends on firms' and individuals' propensity to relocate

with respect to these factors. Also bound up with this is the idea of rationality; that households say will locate in cells in number proportional to those cells' relative attractiveness. There may instead be a probabilistic process operating by which households only tend on the average to locate according to this; a concept not considered.

In addition huge data needs to calibrate the model (with data necessary from several cities in order to give a reasonable cross-sectional calibration), may mean that the model is too limited in application. However, at least there is more theory behind this than behind many other models, and also it attacks the dynamics of the process which Lowry never did.

* * *

Donnelly, T.G. et al. 1966. A Probabilistic Model for Residential Growth.

A model to simulate residential development of undeveloped land.

Components of Model

(1) Attractiveness of land for development: Determined by multiple regression on various factors e.g. access to schools, work, streets, and also by present land values.

(2) Decision process; land chosen for development is selected by random number: The probability of selecting a site is based on attractiveness subject to present and future public policy decisions.

(3) Households are allocated solely according to the density of housing they require. Supply of housing is based on aggregate demand and independent of origins of the households.

(4) Area considered; only the urban foreign areas in which rural to urban land use conversion is considered.

Levels of Aggregation

(1) Spatial: cells of 22.5 acres.

(2) Temporal: generations of three years at a time.

(3) Social: developers and households considered as one behavioral type.

Exogenous Variables

Required are: Site characteristics to determine land attractiveness, land available for development in each cell, public zoning and development policies and density constraints, future demand for housing.

Ability to Forecast

Possible to determine effects of alternative development policies.

Testing Results

Ex post facto testing over the years 1948-1960. Results considered "reasonable," but no statistical test.

Comments

There is no way of determining supply other than from demand. Households are only cursorily considered; with no variations in behaviour with social type allowed, and the effect of their original location on where housing is looked for but not examined. Site selection is based on uniform random numbers, whereas in real life more attractive sites may be selected under a process where probability of being chosen is greater than their relative attractiveness.

See Also

Chapin, F.S. 1965; and Chapin, F.S. and Weiss, S.F. 1962, 1965.

* * *

Drewett, J.R. 1969. A Stochastic Model of the Land Conversion Process, An Interim Report. Regional Studies 3:3:269-280.

A model to simulate changes in land use in an urban area.

Components of Model

A Semi-Markov Chain model of the probabilities of land being converted from one use to another in a given time period, leading to an idea of the amount of time land stays within any particular use, by simulating the results from empire data.

Levels of Aggregation

(1) Spatial: 1 km grid of cells

(2) Temporal: A time period of 1 unit is used which is arbitrary but may be determined from research

(3) Social: is no social grouping.

Exogenous Variables

An empirical distribution of movements from each state to each other state is required, and the probabilities of movement within a time period are derived from here and from assuming an exponential distribution of waiting times before a change of state. Drawing random numbers according to this distribution then allows simulation of the mean waiting time before moving.

Ability to Forecast

This is untried but quite possible by moving the model forwards and

incorporating alternative zoning policies, for instance. And some optimal "minimum cost" transitions might be obtained from examining changes of rent with changes in state.

Testing Results

Not carried out.

Comments

This is an empirical, descriptive model which is not really getting at the processes behind the change, and as such would be difficult to incorporate into a more complex model.

* * *

Engle, R.F. et al. 1972. An Econometric Simulation Model of Intra-Metropolitan Housing Location: Housing, Business, Transportation and Local Government. American Economic Review 62: 87-97.

An econometric model which includes macro-economic behaviour of the aggregate city over time, and also the spatial allocation of the projected activities within the metropolis.

Models

(a) Macroeconomic Model of Industrial Change

This model developes a series of equations for the levels of production in basic and non-basic manufacturing, retail, wholesale, financial and services sector; the sum of which gives the total income accruing to the city. Also derived from here are employment levels and population. These manufacturing levels are derived from information on U.S. and local (Boston) levels of taxation, earned personal income, profits, population and capital investment.

Also derived are expressions for the price indices of services, housing and commodities, and for the average wage rate. This latter then is used to give the income distribution of the population (assuming the income of each occupation is a constant proportion of the average level).

(b) Model of Long-run Change in Population, and Capital Investment in Manufacturing

Population change is due to natural increase (using demographic variables) plus migration (due to economic variables in the U.S. and Boston). Capital investment is based on previous manufacturing growth rates in Boston, and on wages and prices in the U.S. and in Boston.

Thus given present levels of the relevant variables in the U.S. and Boston, population and investment can be predicted and used in the Macroeconomic model.

(c) Model of Spatial Allocation

This is seen as a supply and demand model. Demand depends on site attractiveness, prices and structures available for use. Supply depends on structures available, construction costs, and zoning. A market clearing process then allocates new units.

Testing and Evaluation

Not attempted.

Aggregation

Arbitrary, depending on data availability.

Comments

Linking the macro- and micro-scale processes is useful conceptually, but whether the particular model postulated is any better than others used is questionable.

It has been pointed out for instance in comments in the journal (A.E.R., Vol. 62, p. 101) that the macro-economic model provides a forecast of levels of activities relative to U.S. levels, which assume that Boston copies the U.S.

Also as regards the spatial allocation models, the market clearing process giving a quasi-equilibrium (dependent on constraints) may not be a valid way of representing the real process. It assumes rationality of behaviour by individuals in the land market.

Perhaps the biggest shortcoming is that no explicit forms are hypothesized for the functions in the model. This implies no attempt to get at the behavioural processes which operate in the city. A tremendous amount of work needs to be done before an exact form of the model can be produced which can be tested for its validity.

* * *

Ford, L.R. and Jago, W.H. 1968. An Urban Transportation Simulation presented to the O.R.S.A. National Meeting, San Francisco,

A model of public and private vehicles modelling some cities up to 1990, evaluating costs and benefits of alternate future transportation moves.

The basic algorithm is the '"shortest chain" combines with global iteration, with capacity constraints introduced to speed convergence' (O.R.S.A. Bulletin, 1968, page B-166).

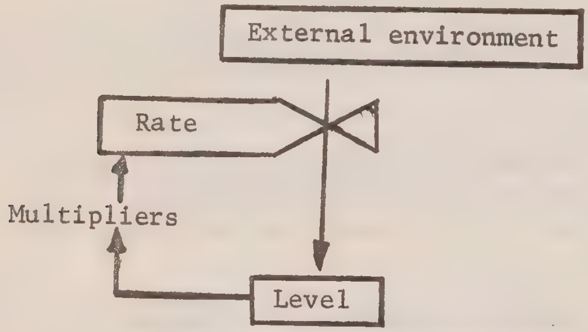
* * *

Forrester, J.W. 1969. Urban Dynamics. Cambridge, Mass.: MIT Press.

A model of the urban system for one city, based on the concept of complex systems. Under this, the model is a system where there is a multiplicity of interacting positive and negative feedback loops linking a large number of system states, whose internal rates of flows are controlled by non-linear relationships.

The Model

The model falls into three parts: the parameter which describe the state of different parts of the system, known as levels; the parameters which control the effect of one level on another, known as rates; and the parameters which relate the system levels to the rates in a feedback loop, the multipliers. Thus in a one state, one feedback loop situation.



Where the rate is represented as a value limiting the effect of the outside world on the system level.

The entire dynamic system is assumed to be closed in the sense that a boundary can be drawn round it such that the external environment can affect system levels, but the system cannot affect the situation in the external environment.

(1) Levels

There are nine principal system levels that represent the state of the urban area. These are: the managerial/professional population, the housing for this group (premium housing), the population of skilled workers, the worker housing, the "underemployed" population (the unemployed and the unskilled), the housing for the underemployed, new businesses, mature businesses, and declining businesses.

The values for these levels are determined by equations of the following form:

Level of variable i at time t = Level of i at (t-1) +

$$+ \Delta t \; \{ \sum_{\substack{j \\ j \neq i}}^{N} \text{Rate of change in variable } j$$

Where N is the number of variables hypothesized to have a direct effect on i.

(2) Rates

These are expressed by equations of the following form:

$$\text{Change in variable } k = \sum_{i}^{m} \text{Value of } i \text{ at previous time period)}*$$

"Normal" rate of change in k* Deviation of the urban system from the normal at time (t-1).

(* = product sign)

Where the set of variables M, which generally includes k, will be those levels whose size is hypothesized to directly affect k.

The normal rate of change is exogenously specified and represents the rate of change in k assumed to occur when the urban system is in its "normal" state.

The deviation of the system from the normal is generally expressed as a form of multiplier which is always greater than zero, and which equals 1.0 when the urban system is in its normal state.

(3) Multipliers

The multipliers are of two forms, both of which are derived from other system variables:

(a) Direct multiplier

Where the multiplier depends just on one system level and which is effectively looked up on a non-linear graph of the value of the variable versus the size of multiplier; where the value of the variable at time (t-1) is given. On such a graph, when the value of the variable is equal to its hypothesized "normal" value then the multiplier will be equal to 1 by definition.

(b) Indirect composite multiplier

Here the deviation of the system from the normal is hypothesized as the joint effect of several aspects of the urban system. Then the value of composite multiplier is itself a product of a set of simple multipliers. Thus the attraction for migration M might depend on the amount of housing H and the jobs J. So the multiplier for M (MM) would depend on both that for H (HM) and that for J (JM):

MM = HM.JM

When the urban system is in a normal state, JM = HM = 1 = MM. In addition it is assumed, by the mathematical operation used, that if there

is either no housing (HM = 0) or no jobs (JM = 0) then MM = 0 and there will be no immigration.

The variables which define multipliers are either system levels, or direct deviations from one or more system levels, at the time (t-1).

Thus levels lead to variables and multipliers all at time (t-1), which in turn lead to rates of change in the period (t-1) to t, giving system levels at time t. This is the essential dynamic structure of the model.

In the model there are altogether 11 secondary as well as 9 principal levels. Examples of secondary levels are levels of taxation and of land occupation. Examples of rates are migration rates, rates of filtering of housing and industry, and birth rates. Multipliers may be rate variables which directly determine rates of change.

Exogenous Variables

(1) Values for system levels at start of simulation.

(2) "Normal" values for system rates.

(3) Hypothesized graph functions for determining multiplier.

(4) Given values for perception time (see below).

Assumptions

(a) Fixed land areas small enough for all people and industries to mix together, so that there is no effect of spatial segregation (Area around is 100,000 acres).

(b) A limitless environment, giving an infinite supply of immigrants and absorption of emigrants.

(c) There is a perception time lag between people inside and outside the system, whereby for new immigrants the attraction of the city as they perceive it at present is the actual attraction in the city X years ago, due to there being a time lag in the diffusion of knowledge.

(d) Attractiveness is always measured relative to the outside environment; not in any absolute terms.

(e) Housing and industry filters through time from higher to lower quality with the longest time being spent in the lowest category and the shortest time in the highest category. Only new businesses premium housing and worker housing are ever built new.

(f) All functional relations in the model are assumed; - there is no attempt to incorporate empirical observations.

(g) There is free mobility in the outside world; as soon as the attraction of the city increases, immigration also increases and continues until

the attractiveness index drops again (lagged by the perception time).

(h) All relations and constants are assumed invariant over the 250 years that the simulation is run.

Evaluation

Given his assumed relationships Forrester runs the model, for the three social groups mentioned with generations of one year in length, until equilibrium is achieved (as the model runs out of land to build on) - taking about 200 years. Then he examines the results against his own (implicit) evaluation of what a city in equilibrium should be like. He finds that there is an excessive percentage of underemployed in the city and in excessive amount of housing relative to industry, leading to a poor tax balance for the city as a whole. He feels improvement of the area to a more dynamic self-supporting city can only be carried out by forcing the underemployed to leave and by replacing the housing of the underemployed with industry. He finds that "social" policies, such as training programmes, and building housing for the poor, only worsen the problem as he sees it, as they increase the attractiveness of the city to the unemployed encouraging further immigration by this class.

Critique

The total novelty of Forrester's approach, together with his very explicit unorthodox opinion on urban revitalization, have caused a great reaction in the form of criticisms and extensions of the model. Criticisms can be found in Fleisher, 1971, Garn, 1971, Horton and Morris, 1970, Kadanov, 1971, Moody, 1970, Newling, 1970, Tobler, 1970, and many other places so just the major criticisms will be summarized here.

It is evident that Forrester has proposed a model, which through its assumptions is very restricted but from which he has tried to draw very general conclusions about the urban area. Some of the assumptions that seem particularly restrictive are summarized below.

(a) Fixed Land Area

Forrester assumes a small area of land in which all economic activities rise and beyond which the city cannot expand, because the mobility equations in the model assume that the different social classes are accustomed to mix with each other. This in itself is a questionable assumption given the real world situation but a further limitation is that the idea of fixed land area essentially closes the system, so that some form of static equilibrium is inevitable in the end. The situation is analogous to population growth models that assume a fixed food supply; and the only reason that Forrester's model does not give a logistic growth curve is because it allows speculation and over-development which initially pushes it above the final equilibrium values.

Essentially the problem is that, as pointed out by Fleisher, 1971, there is no manner in which the computer language used can incorporate

spatial relations and their effects on growth at present. So Forrester has to assume that each small area goes through the same process of land use succession - a concept that has been much questioned by urban writers.

(b) Sensitivity of the Model

Forrester claims that the final results are insensitive to individual parameter changes and so it is unnecessary for him to search for empirical data to validate all his relationships. However other people have tried varying several parameters at once and this has lead to a very marked effect on the model. For instance Horton and Morris, 1971, by changing migration responses to jobs, houses and public expenditive in the city to relations that they felt were more valid, produced a 20 per cent increase in new and mature firms and a 28 per cent drop in unemployment rates of the underemployed. In view of evidence like this one must certainly question Forrester's conclusions about the urban system as a whole.

(c) Policy Proposals

Much criticism has been voiced at Forrester's view of the city as an entity in itself; to be improved irrespection of the results of such an improvement on the underemployed. He makes little real attempt to find an optimal policy or to distinguish between alternative effects of the same policy. For instance a reduction of underemployed housing leads to less underemployed but also higher densities and Forrester makes no attempt to see which effect is more important; whether there is a net benefit as a result of policies or not.

(d) "Counterintuitiveness" in the System

Forrester has often said that one of the major properties of his system is that it is counterintuitive in that it acts in unexpected ways, but many points in the model show that this is by no means entirely correct. For instance a predominance of old buildings in the final equilibrium state is a result of assuming that buildings spend more time as "declining industry" than they do as new and mature businesses, and a similar situation holds for housing. In addition, he assumes that increased density and taxes lead to buildings filtering into the lower class quicker and staying in that class longer. So it is little wonder that in his final equilibrium it is difficult to solve the problem of getting rid of old buildings. Similarly, in dealing with the underemployed, he assumes that declining industry uses more underemployed people, and also that the underemployed migrate to the city twice as fast as any other class when the availability of housing is increased. So again the fact that his final equilibrium contains many underemployed when there is a lot of declining industry; and that his programme for housing the underemployed fails because more enter the city as he builds more housing, is not really surprising. It seems that many of the problems of his city have been assumed into the model in the first place, and with respect to these, the end results do not seem to be counterintuitive.

(e) The Assumption of a Limitless Environment

It is dangerous to put boundaries around something like a city and say that it does not affect the outside world, especially when the city you

are modelling in reality only represents the central city. Even though the rest of the world in total may not be affected by the city, its immediate environment certainly would be. It seems naive to ignore the fact that the structure of the central city will affect the activities located in the suburbs, which in turn will affect the central city again. In addition a limitless environment providing an inexhaustible supply of migrants seems unrealistic in the face of a rapidly declining rural population from which such migrants would have to come. One can certainly conceive of using programmes, of say better housing for the poor, in all cities in a country simultaneously, thus increasing the attractiveness of all of them. In such a situation it seems unlikely that there would be enough people in the rural hinterland to move into all the cities and counteract the effect of more housing being available.

Similarly the effect on a country as a whole of pushing the under-employed out of cities in order to improve living there might well be more detrimental than using policies to absorb such people within the city system. It is not impossible to extend the model to allow for inter-city reactions but it is a little disconcerting that Forrester himself has not attempted to alter or validate his own model but instead has moved on to deal with the world under the same modelling concepts, perhaps because it fits his closed boundary concepts better.

In conclusion it can be said that Forrester made a bold attempt to tackle the city system from a totally new viewpoint. His technical competence seems beyond doubt and his method of attacking social systems may well be useful. However by making little attempt to tie his model in with current urban theory and empirical data he has made it very hard to test the model at anything above an intuitive/descriptive level, and it may need drastic modifications to tie up with current data sources before any real idea of its validity can be gained. Certainly the model should not be ignored, but any policy implications at present coming from the model should be treated with caution. One of its virtues is that public policy can easily be incorporated, but until the model can be validated in its simple form such experimentation is really putting the cart before the horse. The model must be seriously studied to see if it gives us any more insight into the feedbacks of the urban system than the simpler models used at present, as a gain in conceptual and theoretical content does not necessarily lead to a model of greater fidelity.

* * *

Garn, H.A. 1970. An Urban Systems Model: A Critique of Urban Dynamics. Working Paper 113-125. Washington, D.C.: The Urban Institute.

Further simulations with Forrester's model under alternative conditions are carried out, to gain further insight into the model.

* * *

Goldner, W. 1968. Projective Land Use Model: A Model for the Spatial Allocation of Activities and Land Uses in a Metropolitan Region. B.A.T.S.C. Technical Report 219. California: Bay Area Transportation Studies Commission.

A modification of the Bay Area Simulation Study Model; travel times were estimated from the actual network, the gravity function was replaced by an alternative, trips were simulated rather than estimated, vacant land was explicitly included in the model rather than being a residual, capacity constraints were included, and projection was performed on the basis of comparative statistics.

* * *

Graybeal, R.S. 1966. A Simulation Model of Residential Development

New residential land is developed employing user and space interaction, growth and response to policy (land use controls, transportation, etc.). Uses 18 equations; evaluation is carried out.

* * *

Hill, D.M. 1965. A Growth Allocation Model for the Boston Region. Journal American Institute of Planners 31: 111-120.

The "EMPIRIC" Model for allocation of population and employment based on multiple regression.

Model

$$\Delta R_i = \sum_{\substack{j=1 \\ i \neq j}}^{N} a_{ij} \, \Delta R_j = \sum_{k=1}^{M} b_{ik} \, Z_k \text{ or } \Delta Z_k$$

$i,j = 1, \ldots, N$: The predicted variables

$k = 1, \ldots, M$: The independent variables influencing the predicted variables

ΔR_i : Change in i^{th} predicted variable

ΔR_j : Change in j^{th} predicted variable

Z_k : Level of the k^{th} independent variable at the start of the time period.

ΔZ_k : Change in the k^{th} independent variable

a_{ij}, b_{ik} : Regression coefficients.

The model is of the extrapolation form using values calibrated on 1950 data to predict 1960 values.

Predicted variables:

White-, and Blue-collar population, retail, manufacturing and other employment.

Independent variables:

Intensity of land use, zoning practice, accessibility by car and public transit, and quality of water and sewage services.

Evaluation

Reasonable predictions were obtained over a 10 year time period for the Boston area divided into 29 subregions of 100 to 25 square miles in area.

Comments

There is no inherent dynamic property so long term forecasts must be treated with caution; even if, as is suggested, after every 10 years accessibilities and densities are recalculated based on the changes in activity levels predicted for the preceeding 10 years by the model.

It would be difficult to simulate the effect of changing values of independent variables through public policy, as one would have to assume that the values of the regression coefficients remain constant (i.e. there is no feedback from predicted to independent variables).

* * *

Ingram, G.K., J.F. Kain, and J.R. Ginn. 1972. The Detroit Prototype of the NBER Urban Simulation Model. NBER, Columbia University Press, New York, 233 pp.

General Outline

The structure and philosophy of this approach is based in the utility maximizing approaches pioneered by Alonso and Muth. Thus a spatial equilibrium of households is sought based on the amount they are prepared to pay for sites in a competitive framework. The model is simplified from that of Alonso in that only transportation and land costs are considered; the composite good is disregarded. The approach however does relax two major assumptions of the Alonso formulation. First, the model allows for more than one centre of employment, with 14 actually being specified for the Detroit model. Second, it does not attempt to solve for the long run equilibrium in all time periods but only produces an approximation to equilibrium by considering only those households that are actually moving at any time. Thus solutions are always tending towards equilibrium, but are unlikely to ever reach it because of exogenous factors disturbing the system. Thus parameters of the model such as prices are always dependent on current and past conditions.

The structure of the model is a set of seven interlinked sub-models which are solved for each time period operating to locate households into one of 43 residence zones given their work zone.

Components of Model

(1) Employment Location Submodel

This is exogenously specified. It is translated into requirements for different types of worker (classified by workplace), by use of exogenously given distributions of workers for each industry.

(2) Housing Demand

(a) Specification of Movers Submodel

This is based on an exogenous rate of movement of households; modified by the new demands of any industrial centre for worker of this type, and also by increased mobility due to closing down of jobs. Also the characteristics of the households are altered to reflect the change in life cycle, but this is only done to movers and the probability of change is independent of the length of time spent in the current stage of the cycle. Movers are defined according to the location of their job. The movers are further supplemented by the creation of immigrants and new households.

(b) Demand Allocation Submodel

The classification of movers by work sites is converted into demand for residence sites as reflected by prices. This is done in stages. First for each household type at each workplace the households are distributed between housing types according to the relative gross prices of each housing type in the time period. The relative gross prices are a weighted average of the sum of the expected house price by residence zone plus the travel costs of the journey to work to the given work zone. The prices are weighted by the proportion of units of that type available in the zone and by the number of trips to the work place (dependent on income class), and are then added up to give an average expected price for each house type. Thus the importance given to house prices in a given zone depend on the likelihood of choosing a house there by chance, and on the distance of that zone from the workplace.

Given that a housing type has been probabilistically chosen on this basis, the model then needs to find the residence zone that the household will choose. This depends on market supply factors. It may be noted that the assumption is made that house type is always chosen before location, and that the job location is always chosen before housing.

(3) Housing Supply

(a) Vacancy Submodel

This identifies dwelling units vacated by intra-urban migrants. First the proportion of each housing type occupied by the moving workers as classified by work place and income is calculated from exogenously given occupancy rates. Then these are distributed by residence zones in

proportions depending on the number of units available and the number of trips made, as in the demand allocation. For each zone and housing type this is totalled up for all workplaces, and added to it is the number of units left vacant in the previous time (the model attempts always to maintain a level of vacancy or slack in the housing market). Finally for each household type a rate of out migration from the city and a rate of 'death' of households is allowed for and the vacancy totals in each zone increased accordingly.

(b) Filtering Submodel

Filtering will occur upwards or downwards depending on the profitability of a conversion up the quality scale (as measured in the difference in expected prices) relative to the cost of such conversion (exogenously given). The direction of filtering depends on whether the ratio is greater or less than one, and the amount of filtering for any given value of this ratio is exogenously given. Expected house prices vary through zones so filtering changes in the housing stock, which are assumed only to occur in vacant housing in any time period, are calculated for each house type in each zone.

(c) Supply Submodel

New supply of housing is modelled here as including both new housing construction and only form of housing conversion apart from quality change (which is dealt with as filtering). The strategies of supply depend on the demand for each house type in each residence zone, as reflected b expected house prices, and on the cost of the various ways of construction or conversion that can give rise to a given house type. Profit for each strategy is the price of the final house type, less the sum of the price of the initial house type (if any) and the cost of construction. An optimizing model to choose the best strategy could have been adopted but for practical reasons a more *ad hoc* method was used.

In each zone, first the most profitable construction process is used as much as possible, then the second most profitable, and so on, construction being subject to the following constraints.

i) Construction/conversion is profitable

ii) The total construction uses no more than the vacant structures or land available

iii) Zoning constraints (exogenous) are not exceeded

iv) Total construction or conversion in the metropolitan area does not exceed expected demand for any particular housing type (as given by the demand allocation submodel).

(4) Market Clearance Submodel

Given the demand for each housing type, as distributed by work zones, and the supply of housing in each residence zone, the final solution chosen is an allocation of the movers such that future travel costs are minimized.

For any house type h:

$$\underline{\text{MIN}} \quad \sum_i \sum_j \sum_k T_{ijk} \cdot X_{ijk}$$

subject to:

$$\sum_i X_{ijk} = D_{jkh}$$

$$\sum_j \sum_k X_{ijk} = S_{ih}$$

T_{ijk} = cost of trips from i to j for household type k

X_{ijk} = number of households type k, working at j who are allocated to i

D_{jkh} = demand for house type h by household type k, working at j (as given by demand allocation)

S_{ih} = supply of housing of type h in zone i (as given by market supply models)

$i = 1, \ldots, n$ = residence zones

$j = 1, \ldots, m$ = work zones

$k = 1, \ldots, Y$ = incomes of households

$h = 1, \ldots, H$ = house types.

To solve this model supply must equal demand and if it does not then dummy housing units or household units must be added as appropriate until this equality is met. When supply is in excess, the location of the dummy households may readily be interpreted as reflecting marginal land where the shadow prices are zero. However in the case of excess demand the interpretation put on the dummy housing units are that they represent hotels in which the unsuccessful bidders for housing temporarily stay, or alternatively they represent households who move out of the city for a while and then return in the next time period to rebid for housing. Both of these seem unrealistic interpretations and in the model it seems that mobility and outmigration rates do not depend on housing price changes as might be expected. This weakness in the short term dynamics of the model is recognized by the authors who stress the difficulty of dealing with large short term changes in demand. This may be seen as a major weakness in the model.

(5) Price Changes

The dual of the assignment problem solved for household allocations gives the shadow prices of the different residence zones for each house type. These may be interpreted as location rents, and in fact for the time period in question the 'equilibrium' prices generated for the current period are equal to the sum of these rents and of the cost of the least expensive way of supplying that housing on the marginal land (where location rent is zero). This latter provides the absolute basis upon which the price variations, as determined by locational advantages with respect to all work sites are imposed.

Given these current equilibrium prices these are used to alter the expected house prices that were entered into the model at the start of the time period, so that expected prices in the next period partially reflect current supply and demand in the housing market.

$$EP_{t+1} = a \cdot EP_t + (1-a) \cdot CP_t$$

a is an exogenously given constant $0 \leq a \leq 1$

EP_t expected house price at time t.

CP_t current equilibrium price at time t.

The averaging of current and past prices to give the future price is thus carried out in an adaptive expectation framework. This is the

main internal variable that changes for successive time periods, and it is an attractive feature of the model.

Level of Aggregation

Time: arbitrary, but periods of approximately one year are suggested.

Space: aggregations of census tracts into zones varying from 1 to 250 square miles. They seem to be defined more by population than size.

Social: households are broken into 72 categories, classified by income, age, and education of the household head, and by family size. For calculating mobility characteristics in the model all these are used, but final demand is assumed to vary only by income. Housing is divided into 27 categories.

Exogenous Variables

The major arbitrary exogenous variables are the change in employment, the response function of filtering with respect to expected profitability, and the value of the averaging constant in calculating future prices. All mobility rates and worker characteristics by industry are also exogenous but are estimated from Detroit and San Francisco interview data. They are assumed constant over time. In addition, initial house prices are exogenous but these are internally generated thereafter.

Testing of Model

No empirical testing has been carried out. Although Detroit was taken as a basis for the running of the simulations to see if the results were reasonable this was purely an illustrative case as data for all the parameter were not available for Detroit. The model has since been applied to Pittsburgh where there is a more complete data set, but again results have been evaluated largely as regards their basic plausibility rather than by rigorous testing.

Comments

The model provides a compromise between the theoretical structure of residential equilibrium in the housing market and the practical needs of developing a tractable model. The theoretical side has been more rigidly specified than in previous models of this type and in this sense it may be seen as an advance.

However it only models the housing market, and only the purely economic aspects of that, as the authors recognize. They anticipate eventually including non-economic aspects such as racial segregation and other market externalities. They also hope to include basic employment as an endogenous component in the model, but this is only one of the many other facets of the city that the model would have to relate to if it is to adequately simulate urban growth. In particular the mobility parameters should not be made insensitive to many conditions and opportunities in other parts of the urban economy and in the outside world.

If such interlinkages could be specified the model might well be very useful, providing of course that the optimizing framework of household equilibrium gives a reasonable fit to the real world situation. It is good to see the spatial aspects of the problem more adequately dealt with than in comparable studies such as the San Francisco urban renewal model.

* * *

Irwin, N.A. Review of existing land use forecasting techniques. Highway Review Record, No. 88

Discusses:

(a) The Penn-Jersey Transportation Study

Use of linear programming to simulate residential development under maximization of aggregate rent-paying ability subject to constraints. Utilizes household aging, migration, income changes, and transportation costs under extensive disaggregation. Data needs are great.

(b) RAND Model

Built to study the transportation-land use linkage. Carried out for six month generations. No calibration or testing carried out.

* * *

Kadanoff, L.P. and H. Weinblatt. 1972. Public policy conclusions from urban growth models.

Modifications of the Forrester model are used to show how Forrester's assumptions to an extent determine his public policy conclusions.

Models

(a) Many-city Forrester Model.

A series of cities of the Forrester type are postulated with the cities linked by migration. Migration from city 1 to city 2 is seen as: proportional to the population of each city, proportional to the attractiveness of city 2 as perceived in city 1; and inversely proportional to the attraction of city 1 to the migrants. This inter-urban model represents a larger proportion of the nation and is used to examine the effect of applying a public programme to all cities. The attractiveness and proportion of the population in suburbs and rural areas are assumed constant (as in the original Forrester model).

(b) The National Metropolitan Model.

Here suburbs are included explicitly as part of the model, with the same type of dynamics as the central city. They are assumed to have no limit to their area. Migration between city-centre and suburb, and between different cities is guided by the same equations as in the many-city model. The rural area is assumed to have fixed attractiveness.

In addition, instead of assuming all production by industries is always used, a demand side of the function, determining the growth in industry, is also included. Demand is a function of total population in a city, and is divided into demand for basic and non-basic industry. The production to meet this demand is then divided between the cities on the basis of relative sizes in each city of the enterprise multiplier, the amount of industry already present, and the ease of travel between the area of demand and that of production.

Workers can live in one part of a city and work in another. The city could potentially be split up into as many sub-areas as required.

Results

When Forrester's public programmes are applied to these models, they gave different results to those of Forrester's original model. For instance demolishing housing of the underemployed, which led to a fall in the numbers of these people according to Forrester, now has no such effect. This is because the new industry which the extra space would attract already exists in the suburbs of the cities, so that there is no effective increase in industrial activity. On the other hand, programmes that give training to the underemployed, and those that give new jobs to the skilled workers, which had little effect for Forrester, work far better here.

In addition they propose a new index of well-being in a city which was included in the Forrester model but not used by him for this purpose; the level of the attraction of the cities to incoming migrants.

Evaluation

Still no attempt is made to relate the model to any empirical data or relationships. However they state that this will be done.

Aggregation

Spatial: Each city is split into just two regions.

Temporal: arbitrary

Social: There are just three types of people, housing, and industry, as for the Forrester model.

Comments

An interesting attempt to introduce the dynamics of inter-urban relationships. However the validity of the model stands or falls on the validity of the formula for interaction between cities.

Also many of the shortcomings of Forrester still apply; such as the lack of relation to empirical data and urban theory. There is also the problem that suburbs of unlimited size violate the assumption of Forrester; as his model applies to small areas where social groups are mixed rather than segregated.

* * *

Kain, J.F. 1964. The Development of Urban Transportation Models. Papers and Proceedings of the Regional Science Association 14: 147-173.

Model

A description of a model developed by RAND to study Urban Transportation. The Model is essentially static; but sub-models, for use as inputs to the transportation study, have been developed for the change in location of employment and for household behaviour.

(a) Employment change:

$$\frac{W_j^k}{W^k} = f(X_j^1, X_j^2, \ldots, X_j^n)$$

W_j^k = Employment in area j in industry k

W^k = Total employment in industry k

X_j^i = i^{th} independent variable (describing access, etc.)

Simulations were carried out for alternative time-paths of independent variables and parameter changes.

(b) Household behaviour:

$$\frac{W^k_{ij}}{W^k_j} = f'(X_1, X_2, \ldots, X_n)$$

W^k_{ij} = Number of employees working in area j who live in area i

W^k_j = Total number of employees in area j

X_i = i^{th} independent variable (describing journey to work time and cost)

Simulations were used to describe the predicted residential pattern of non-white workers, given their work places, under conditions of no segregation, as opposed to real world patterns occuring under segregation.

Comments

A large number of exogenous variables are necessary. There is no theory provided to describe the form of the functions of the independent variables. No real evaluation is provided as the work is described as only half-completed. Other sub-models of the urban system are hinted at in the paper, but not described.

Krisbergh, H. and Nelson, C.W. 1972. A Search Procedure for Policy Oriented Simulations: Applications to Urban Dynamics

Method

Given as an output of a simulation at any time t as a state trajectory

$$\underline{X}(t) = (X_1(t), X_2(t), \ldots, X_n(t))$$

where $X_i(t)$, i=1, . . ., n is a vector of state values

And given a control trajectory of values of control variables in the system on which the state values depend

$$\underline{\emptyset}\ (t) = (\emptyset_1(t), \emptyset_2(t), \ldots, \emptyset_r(t)$$

where $\emptyset_j(t)$ is a vector of control values (j=1,. . .,r)

Then it is desirable to be able to choose $\underline{\emptyset}(t)$ so as to give a pre-determined desirable output $\underline{X}(t)$. To do this it is necessary to state what desirable values of $\underline{X}(t)$ are. Then an objective function of the following form may be set up:

Choose $\emptyset(t)$ to:

$$\text{Min} \quad U(\emptyset,t) = \sum_{i=0}^{T} \sum_{i=1}^{n} \{W_i^u(t) \left| X_i(\emptyset,t) - R_i^u(t) \right| + W_i(t) \left| X_i(\emptyset,t) - R_i(t) \right| \}$$

$R_i^u(t)$ = desirable upper bound on X_i at time t

$R_i^\ell(t)$ = desirable lower bound on X_i at time t

$W_i^u(t)$ = weighting factor for the deviation from the upper bound

$W_i^\ell(t)$ = weighting factor for the deviation from the lower bound.

The weights are chosen first of all to ensure that all X_i are measured in similar magnitudes, and second to rank the X_i in order of importance in satisfying the criteria. Then the model proceeds by using a version of pattern search given some arbitrary initial values of $\emptyset(t)$ until this breaks down. Then a new starting point is chosen randomly and pattern search applied from here until it also fails. Then given these two values as reference points a new pattern search in the optimal direction is chosen with respect to these. This is repeated until one of several possible terminating criteria is reached.

Before running the following parameters must be specified:

(a) The variables in the control vectors

(b) The permissable values that these controls or policies can take on, with a penalty being imposed for going outside these.

(c) The minimum desirable value of the objective function.

(d) Limit on the number of runs of the simulation model.

(e) Limit on computer time.

(f) Minimum increment in control values used in the search.

The model runs through cycles of alternately choosing $\emptyset(t)$ and evaluating $U(\emptyset,t)$ until one of the limits (c), (d), (e), (f) is reached.

Application to Urban Dynamics

The method is applied to Forrester's urban model as an example.

(a) Objective function:

Low unemployment for the under employed (approximately one person per job).
An underemployed to housing available ratio of about one (implying excess housing is destroyed).

A high upward social mobility for the underemployed.

Three policies are tried:

(i) Slum housing elimination and new industry construction are tried together at constant values for all t. The results come out as almost identical to Forrester for this problem.

(ii) The above two policies plus one of training the underemployed are tried together, each being allowed to take on three different values during the simulation run. The result is that unemployment falls faster than in (i), with the other two objectives behaving much the same. This indicates that job training does work contrary to Forrester's beliefs.

(iii) Three other programmes are added in: construction of high income housing, job creation for the underemployed, and demolition of declining industry. All six programmes were held fixed for the entire time period. In this situation there is an even sharper initial drop in unemployment, and also for a short time social mobility reaches very high figures. Again results are contrary to Forrester's belief that four of these six programmes are no good.

Comments

This method provides an efficient method of evaluating combinations of programmes in a way which Forrester was never able to do, thus allowing for something of a more objective evaluation of the results of a simulation. The efficiency is indicated in that none of the above programmes took as much as 60 seconds on a UNIVAC 1108. To work the method, some form of optimal objective function is necessary and this is often difficult to determine. However the property of the method of allowing interaction between model and investigator is very useful, and it seems to fill a crucial gap in the process of model evaluation.

* * *

Lamb, D.R. 1972. Research of Existing Land Use Models. Pittsburgh: S.W. Pennsylvania Regional Planning Commission

Discusses:

(a) Philadelphia Activities Allocation Model

Consists of six sub-models: locating residences, manufacturing, and non manufacturing; and modelling space consumption by streets, residences and manufacturing. It is an econometric model using six multiple regression equations to allocate growth to 192 regions in five year generations.

(b) The Connecticut Land Use Model

Allocating growth to 169 towns in Connecticut using nine simultaneous equations to describe shifts in employment and population. Ten year projections are made.

Lathrop, G.T. and Hamburg, J.R. 1965. An Opportunity-Accessibility Model for Allocating Growth. Journal of the American Institute of Planners 31: 95-103.

A model for allocating growth to areas sequentially based on their accessibility to the centre of development in a region.

Model

$$A_j = A\ [\exp\{-1.0\}-\exp\{-1(0+0_j)\}]$$

A_j = Amount of activity allocated to j

A = Total amount to be allocated

1 = probability of a unit of activity being sited at a given opportunity

0 = Opportunity for siting a unit of activity in the zone preceding j

0_j = Opportunity in zone j

Where zone j is the next most accessible zone to the centre after "the zone preceding zone j".
So development is allocated to zones in order of decreasing accessibility.
Opportunity is defined as the product of land available and the density of activity on that land.

Then as the probability, 1, falls; the development can be seen to be more spread out.

Comments

The model is very simple with development based on access primarily, but the model is easily manipulated and run for alternative zoning laws and density controls.

* * *

McLoughlin, J.B. 1969. Simulation for Beginners: The Planting of a Sub-Regional Model System. Regional Studies 3:3:313-323.

A model to simulate the development of a sub-region.

Components of Model

(1) Population: Linear extrapolation from population and employment at previous time period.

(2) Employment allocation ("basic" industries)
Based on policy for development of sub-areas.

(3) Population allocation to cells (approximate)

Based on employment data and density of workers per acre; giving an idea of the land filled by people.

(4) Population allocation to cells (detailed).
Based on formula for dwelling condition relative to previous time period, and on subjective judgement on clearance, renewal and building of housing. Any people not accommodated in the cells when these variables have been considered are allocated elsewhere based on access to facilities and land quality, giving rise to a modification of the cells filled by people.

(5) Service and Labour Oriented Industry Location
Industry and hence employment are allocated according to population distribution, and these generate further employment, and population to be settled giving a final idea of amount of land filled.

(6) Traffic
Movement for three types of trip is estimated and the need for new routes evaluated.

(7) Shopping Centres
Given values of expenditure for the population by cells, these are allocated subjectively according to planning policies, and their size is calculated from a gravity formulation.

Levels of Aggregation

(1) Spatial: 3 levels, 9 subregions, split into 98 areas for trip generation, and 600, 2 km. square areas for population and service employment allocation.

(2) Temporal: One generation is five years after which everything is adjusted to give a market equilibrium.

(3) Social: Consider just 2 types of worker and 5 types of industry exist.

Exogenous Variables

Population forecasts by subregion, modified by employment forecasts if there is any discrepancy, are needed. For the whole area aggregate information is necessary on children of school age, number of households, household expenditure and car ownership.

Ability to Forecast

Is oriented to this purpose based on policy decisions derived from discussions with local planners.

Testing Results

Very difficult because of the large amount of subjective judgement, and as a result this has not yet been done.

Comments

It uses a large amount of subjective judgement derived from discussions with planners about their own decisions. A useful approach because very often

the planners are the real decision makers; but because of this the model has little generality outside the areas of its use (Leicester, England), as no attempt was made to generalize on how planners in general behave in determining policies. A less subjective model might be possible without losing too much accuracy (if indeed the model has any accuracy).

See Also

McLoughlin, J.B., ed. 1969. Leicester and Leicestershire Sub Regional Planning Study, Vols. 1, 2, Leicester.

* * *

Malm, R., et al. 1966. Approaches to Simulations of Urban Growth. Geografiska Annaler 46 Series B: 9-22.

A model to simulate the growth of apartments.

Components of Model

A model to simulate apartment development.

Probability of development is inversely proportional to the cost of building on a site and to the cost of extending services to the site; and is proportional to the ease of building on the soil type present.

Actual developments are determined by drawing random numbers.

Aggregation Levels

(1) Spatial: 400 in X 400 in cells

(2) Temporal: not specified

(3) Social: all buildings are of the same type.

Exogenous Variables

Data on building and services costs, soil type, and projected amount of development are required.

Comments

It is difficult to evaluate performance based on any criterion other than goodness of fit to real world, because of its simplicity.

* * *

Massie, R.W. 1969. A System of Linked Models for Forecasting Urban Residential Growth. Master of Regional Planning Thesis, University of North Carolina.

A behavioural approach to the simulation of land conversion from rural to residential use.

Components of Model

(1) Landowners: Probability of selling land to developers is based on personal characteristics of the landowners.

(2) Developers: Probability of buying and subdividing land into certain development types is based on access qualities of the land, and on personal characteristics of the developers.

(3) Households: Probability of buying a house is based on personal characteristics.

(4) The above components are calibrated by multivariate discriminant analysis on a series of independent variables. The probabilities associated with each group are then calculated on the basis of the discriminent scores on the independent variables and the characteristics of the individuals considered, and an output of these probabilities is the result of the model.

Aggregation Levels

(1) Spatial: Cells of 22.5 acres

(2) Temporal: Not specified

(3) Social: Only one household type is considered and a discriminent function is calculated on the basis of this one group. Similarly only one landowner and one developer type are assumed.

Exogenous Variables

Variables to measure the characteristics of landowners, together with information on whether these landowners have sold their land, is necessary to calculate the discriminant function, and similar data needs are necessary for the developer and household. Also, information is needed on the growth in demand, which can then be allocated to cells on the basis of their probability of development in the simulation, and on the actual characteristics of landowners, developers and households in the future.

Testing Results

The only results tested were the efficiency of the discriminant functions in separating developed from undeveloped land. This was "moderately" successful, varying from 41% - 80% correct classification.

Comments

Is really a form of comparative statics model based on exogenous projections of demand. The locational effects on the choice of housing by households is ignored, which may explain the very low discrimination achieved in this group (41% correct classification).

The effects of supply on demand and vice versa are not considered. Detailed information on the character of actors in the model in future time periods, from which the discriminant scores would calculate the probability of a landowner, say, selling his land, would be difficult to obtain.

* * *

Mesarovic, M.D. and A. Reisman , eds. 1972. Systems Approach and the City. Amsterdam: North Holland Pub. Co.

This book is a collection of papers presented at the 5th Systems Symposium, Case Western Reserve University, November 1970.

Mesarovic: "Introduction" (pp.1-7)

He describes the systems approach used here as characterized by:

(a) A consideration of problems from the broadest possible view: each system is the sub-system of a greater one.

(b) An emphasis on relations between variables and inner causal mechanisms, making better prediction possible than through trend extrapolation.

(c) Providing a common language, in decision making and information processing terms, so that different types of systems may be compared.

(d) A combination of the above three elements to provide an integrated truly dynamic system which may be studied by simulation and other quantitative techniques.

C.A. Doxiadis:"Ekistics" (pp.8-42)

Explains the role of ekistics in providing a common language for studying settlements and defining within them the role of man interacting with his environment. He argues that this leads to ways of measuring the quality of life and thence to ways of improving it.

A. Alfield and D. Meadows: "A Systems Approach to Urban Revival" (pp.43-67

Describes the role of the Forrester model in finding the "best" methods of urban renewal. The basic principle is that net attractiveness of the city with respect to the rest of the world is never high for very long since beneficial programmes are counteracted by negative feedback.

(The logical counter-argument is that cities will never be much worse than the environment either as bad policies lead also to negative feedbacks; meaning that the only real problem is to improve the environment as a whole.)

B. Harris: "Change and Equilibrium in the Urban System" (pp.68-86)

Divides urban models into equilibrium ones that ignore dynamics, and dynamic ones that are not related to any particular equilibrium and thus make it difficult to determine the path to be taken to any specified equilibrium. He relates the two approaches through the notion of negative feedback which makes a dynamic model an equilibrium seeking one.

He also discusses intuitively problems of high collinearity that tend to occur in solutions near the equilibrium, making accurate statistical estimation very difficult.

E.S. Savas: "The City as a System: A Political Administrative View" (pp.87-96)

J. Richardson and T. Pelosci: "A Multi-Level Approach and the City" (pp.97-131)

E. Cusken: "Some Afterthoughts on Four Urban Systems Studies" (pp. 132-155)

These three papers look at the city through the role of government agencies. Savas proposes that this role may be better analyzed if described in systems theoretic terms. Richardson and Pelosci go beyond here to suggest that the best way is to view government as a hierarchical multi-level system which controls the environment based on its observations of it and on certain goals. Cushen records the results of four systems studies applied in certain cities to solve particular problems, these being part of a larger study on the utility of such approaches in this context.

D. Drew: "Traffic Application of Multi-level Systems Theory" (pp. 156-175)

R.L. Meier: "Evolution of Minibus Systems" (pp. 176-189)

R. Nelson: "The Northeast Corridor Transportation Study" (pp. 190-207)

These all focus on transportation. Drew discusses a hierarchical control system that may be set up to minimize congestion and maximize efficiency on highways. The hierarchy runs from individual control elements on each ramp up to an overall control mechanism. Meier describes the structure of the Hong Kong minibus system and its prospects in the face of the onset of mass transit. Nelson talks about the experiences of the NE Corridor Transportation Project in the U.S.; how it achieved compromise between the ideals of the systems and transportation engineers, and the practicalities of data availability, forecasting, and the difficulties of trying to incorporate all the impacts of transportation.

A. Blumstein and R. Larson: "Models of a Total Criminal Justice System" (pp. 208-252)

A. Blumstein: "Systems Analysis of Crime Control and the Criminal Justice System" (pp. 253-273)

I. Balbus: "Urban Ghetto Revolts and Local Cricriminal Court Systems" (pp. 274-313)

These all concentrate on the criminal system in the city. Blumstein and Larson set up a systems model to simulate the flow of criminals through the system, the costs involved at each step, and feed-back with the environment of the criminal world in terms of the probabilities of rearrest and of switching crimes. Blumstein describes an alternative model which is of the interactive type, allowing actors in the system to simulate alternative policies and immediately see the results. Balbus compares the reaction of the courts to ghetto riots in three different cities and shows that it is a function of the form of the political system and of the size of the riot. He argues that because of this the crime system cannot be seen as closed since the size and type of crime partly determines the type of justice handed down.

E. Savas: "A Computer Based System for Forming Efficient Election Districts" (pp. 314-342)

He outlines an electoral districting algorithm to equalize the population of voters in each district so that equal access to polling machines is possible for everyone.

A.L. Service, et al.: "Systems Analysis and Social Welfare Planning: A Case Study" (pp. 343-374)

A systems model is constructed of the set of community organizations in a Jewish community in Cleveland and, subject to certain goals, an optimal allocation of money is achieved between the different organizations according to the output of each (where output is measured as the quantity, quality, and relative importance of the work done), such that total output of the system (the sum of individual outputs) is maximized.

S. Goldstone: "Dilemmas of Systems Analysis of Urban Public Programs" (pp. 375-388)

The major theme of this discursive paper is that the application of systems analysis to improving public programmes must beware of the dilemma between producing an optimal solution and satisfying political constraints in the system.

S. Milgram: "The Experience of Living in Cities" (pp. 389-417)

Milgram argues that city life starts with external facts of the city (such as large numbers and high density) which lead, through adaptive mechanisms within the individual, to the typical "tone and behaviours of city life." This will vary between cities as city characteristics vary.

C. Flagle: "Health Systems: An Urban View" (pp. 418-430)

Three models are described of health systems which discuss the operation of the system given a certain population structure. Emphasis is placed on the problems of the spatial structure of the systems and of demand forecasting as major areas of future work.

F. F. Gorschboth: "Systems Analysis of Urban Air Pollution" (pp. 431-450)

Y.Y. Haimes: "Pollution and Ecology" (pp. 451-460)

The last two papers concern themselves with pollution. Gorschboth describes a systems model of air pollution given the location of emissions and physical and climatological conditions, and builds in control mechanisms with level of operations dependent on specified goals. The controls respond to, and eventually will anticipate, changes in weather conditions. Haimes reviews work on systems analysis as a way of defining pollution problems, with emphasis on water pollution.

Comments

A very wide range of partial problems are tackled here, but given the emphasis of the introduction on the utility of integrating the systems the lack of any overall view is disappointing. The impression left is one of a series of relatively trivial problems with systems techniques used as a method of problem description as much as a method of solution. The real insight that may be gained from using systems approaches to integrate very different problems has not been utilized. Underemphasis on the forecasting problems of dynamic models and on the spatial elements that are important in city structure (mentioned only by Harris, Drew, Flagle and Gorschboth) is also evident.

* * *

Morrill, R.L. 1965. Expansion of the Urban Fringe: A Simulation Experiment. Papers and Proceedings of the Regional Science Association 15: 185-199.

A model to simulate development on the urban fringe.

Components of Model

(1) Selection of land for development: by choosing sites randomly from the map, rather than a choice based on distance from present development.

(2) Probability of a site being developed: Based on combination of values of access to major development centres (schools, work, shops, etc.), and distance to major roads. This is modified by quality of site for building.

(3) Kind of development: Probabilities of various types depend on zoning and empirical distributions.

(4) Size of development: Arbitrary probability of 0.5 of developing the whole site, and of 0.5 of selling it off as individual lots.

(5) Density of development: Based on empirical distributions.

Levels of Aggregation

(1) Spatial: individual land holdings are considered.

(2) Temporal: variable, depending on the amount of development specified exogenously.

(3) Social: people not considered; development is divided into businesses, apartments and houses.

Exogenous Variables

Empirical distributions of development at various distances from "development centres", of the proportions of different types of development and of the density of development of different types are all required.

Ability to Forecast

This is possible by iteration under alternative policies.

Testing of Results

Carried out by comparison of actual data and modelled data for 1964, based on original data for 1957. The data compared statistically were frequency distribution of sizes of developments and of distances of these from development centres. However, this comparison only shows that there is a consistency in the distributions over seven years, rather than there is any real theoretical validity for the model. As these distributions are fed into the model in 1957, it is only natural that the model should reproduce the same frequency distributions.

Comments

The model is based on empirical probabilities which are assumed constant over the simulated period. Here, rather than producing a model of thoretical content, there is a danger of "reproducing our own ignorance" (Garrison, 1962).

The model assumes that different types of building have an equal chance of development anywhere within a particular zoning area, even though overall probabilities of a site being developed do vary.

No real account is taken of the effect of individual behaviour on the land market either in generating overall supply and demand or in determining where particular types of development are most likely to occur. Instead of this some overall probability, based on a composite access function to a site, is assumed.

* * *

Morrill, R.L. 1965. Migration and the Spread and Growth of Urban Settlement. Lund Studies in Geography, No. 26B, Gleerup. Lund.

Models to simulate the growth of a central place system.

Simple Model

Area growth and town development based on migration through a single centre, and town growth according to arbitrary size and spacing rules based on central place theory. For illustrative purposes mainly; little practical use.

Experimental, Detailed Model

Components

(1) Industry: all assumed to be one type, with probability of location dependent on size of urban population and existence of transport (railway) through the area. The relative weightings of this factor were subjectively determined.

(2) Transport routes: rail only considered; roads were considered "too complex." Probability of location proportional to population and inversely proportional to distance and construction costs. Again the

weighting of factors was subjective. The set of likely routes to be considered was also subjectively chosen.

(3) Central Place Activities: Threshold populations were picked based on empirical data about the region, and centres were picked which had sufficient population within their market areas. Those potential sites chosen were based probabilities dependent on population and the existence of rail links.

(4) Migration: Probability of migration is proportional to the population of the smaller area with urban population counted twice and inversely proportional to distance, raised to an arbitrary exponent. The number of individuals migrating from each area was taken to be a given proportion of the population of the area multiplied by the total probability of migration from that area to all other areas.

In all cases probabilities were determined by drawing from a uniform distrition of random numbers. The various components were added into the model sequentially at each generation which reduced the interdependence between them that exists in the real world where they are all occurring over the same period of time.

Levels of Aggregation

(1) Spatial: units used were Swedish parishes and population 100-5000 and size some 25 km^2

(2) Temporal: generations of length 20 years were used

(3) Social: individuals were used but not divided by social characteristics.

Exogenous Variables

Rates of population growth, estimates of central place thresholds, total growth in urban population and resultant demand for industry of all types are required. Also the many subjectively determined weightings are effectively exogenous to the model.

Ability to Forecast

This is possible for migration, subject to assuming that given conditions hold for settlement growth. More difficult to predict the patterns of settlement especially for the areas of lower population because the model is sensitive to and consolidates on random decisions which may easily be different in the real world.

Testing Results

Carried out for areas of semi-industrial Sweden, 1860-1960. Chisquare and Kolmogorov-Smirnov tests on the aspatial frequency result of the various components of the models were fairly successful. Tests of the two dimensional patterns were only subjective: noting similarities in the process of development. Also nearest neighbour tests were compared; carried out for similar sized towns.

Comment

Relations were very simplistic especially for industrial location components and there was a high degree of dependence on subjective values and empirical evidence in calibrating the model. The result was a model based on empirical constants rather than on inherent processes operating within the cities, which might generate growth parameters as internal to the model. The high variability between results of different runs of the model suggested by Morrill could well be a result of the model being too generalized and being a case of trying too much at once.

Also the set up of the model seems to be for only those types of landscapes of fairly even settlement that have undergone gradual but widespread industrialization over time, thus allowing the principles of central place theory to come into action as a dominant factor.

However, valuable insight is given into the role of chance in the historical development of the landscape.

The model is summarized in:

Morrill, R.L. 1963. The Development and Spatial Distribution of Towns in Sweden. Annals of the Association of American Geographers 53.

Morrill, R.L. 1962. Simulation of Central Place Patterns over Time. In K. Norborg, ed. The I.G.U. Symposium on Urban Geography. Lund Studies in Geography, Series B, No. 24. 109-120, Lund, Sweden.

Morrill, R.L. 1965. The Negro Ghetto: Problems and Alternatives. Geographical Review 55: 339-361.

A simulation of the spread of the ghetto in Seattle.

Model

The model assumes contagious diffusion is a reasonable way of representing the spread process. Natural increase and the number of immigrants are determined externally to the model, with immigrants' locations in cells probabilistically dependent on the number of negroes in those cells. The number of negroes moving is also determined exogenously. The probabilities of their moving to certain cells depend on a 9 x 9 mean information field. If the destination is in a negro area the move is immediate, if the destination is a cell with no negroes in it, this cell has to be connected twice before a move is made. There is a limit on the maximum number of families in each cell, determined exogenously.

Evaluation

Qualitative: Morrill states that the same type of process in intensity and rate of expansion (the number of new cells containing negroes given the exogenously determined total number of negroes moving). However he notes that the direction of spread tends to be systematically different and suggest

that a reason for this is that present house values were not taken into account (it would be more difficult for negroes to move into rich areas).

Comments

The probability assignments and degree of resistence to ghetto expansion seem arbitrary, with no theoretical or empirical rationale for the values chosen. The model assumes all areas are equally resistant to expansion and that as soon as there is even one negro present, resistance to any other negroes moving here is zero. This is very simple approach to a complex phenomenon.

Pilgrim, B. 1969. Choice of House in a New Town. Regional Studies 3:3: 325-330.

A model to simulate selection of a new house by a household.

Components of Model

(1) Allocation of households to new houses, a gravity formulation:

Allocation of i to type k = $N_i . N_k . A_{ki} . C_i . B_k$

N_i = number of household type i

N_k = number of house type k

A_{ki} = attractiveness of house type k to household type i

C_i, B_k = constants which can be determined

(2) Attractiveness:

$$A_{ki} = \frac{1}{D_{ki}}$$

D_{ki} = difficulty of matching type i to type k
$= f(d_1) + f(d_2) + f(d_3)$

$f(d_1)$ = Function representing change in D_{ki} with change in rent

$f(d_2)$ = Function representing change in D_{ki} with change in overcrowding (house size)

$f(d_3)$ = Binary function representing relative attraction of flats versus houses

Levels of Aggregation

(1) Aspatial

(2) Temporal: one generation represents one year

(3) Social: households divided by age, size, income

Exogenous Variables

Required are projected numbers of houses and households of each type, and empirical formulations of the difficulty functions. Employment and population forecasts are envisaged as being linked to give household projections.

Ability to Forecast

Can test alternative housing programs, changes in rent structure, effects of job opportunities on town growth. It has the ability that you can intervene at any point to adjust housing development plans depending on which types are in most demand for those attracted to the region by the new employment.

Testing

Is untested as yet.

Comments

The model is aspatial, giving no idea of where to build. There is no probabilistic input; the model assumes each household types acts exactly according to the preference functions calculated for it, with no individual variation. Supply of housing is determined only by demand for it. Parameters are estimated for a single point in the past where there is information on the number of households in each type of housing, with extrapolation from here. There is no attempt to look for a theoretical rationale for the existence and form of these functions.

* * *

Ross, P.S. and Partners. 1972. Promus: A Simulation System for Urban Analysis and Management. Toronto.

Model

(a) Financial Policy Planning Subsystem

The purpose of this is to state the expenditure requirements, and revenues accruing to the city as a result of any policy and the effect of this on the community over a period of time. This will allow the evaluation of the results of alternative fiscal and planning policies.

Sub Models

(1) Policy Implementation Model - specifies the impact on the community of each possible programme.
(2) Programme Performance Model - states the results of the programme over time.
(3) Budgetary Expense Planning Model - states the necessary outlays.

(4) Revenue Forecasting Model.

(5) Financial Cash Flow Model - balances the budget and states benefits, deficits.
(6) Policy Decision Model - evaluates alternative policies and their impact on the budget.

Thus community data is converted into budget and cash flow forecasts.

(b) Community Model Sub-System

The community is divided into neighbourhoods with specified attributes. The relation of these attributes to determining variables is through multiple regression models. In this way the substantive effect of exogenous inputs such as financial policies are evaluated.

Sub-Models

(1) Time Oriented Metropolitan Model - This distributes basic and non-basic industries and population, given exogenous input of basic industry and travel-time distributions. A small area sub model distributes this to communities in amounts consistent with the macro forecast, splitting the land into five types of use:
Unusable, basic industry, non-basic industry, residential, land used by public sector.
The outputs of the model are: land use and households by small area; total employment.

(2) The Neighbourhood Model - On the basis of the T.O.M.M. output, this estimates other attributes of each small area, taking into account both present values, and the change in values from the previous time period.
The output is in terms of sociological variables under five categories:
Housing, education, employment, health, welfare.

(3) Population and Income Distribution Models - Given estimates from the two previous models, this distributes population by sex, race, age, occupation, and housing occupied. Then the distribution of income earned is derived.

(c) Linkage Matrices Between the Two Models

These are based on intensive examination of the present structure of local government to determine who was responsible for the different types of projects and the nature of the decision making.

Then each programme was examined in detail and its characteristics put under four headings:

(1) Observable Output (impact on community)

(2) Exogenous Input (areas influencing the project over which local government has no control)

(3) Controllable Input (decisions about executing and changing the programme over which there is direct control)

(4) Decision Rule (an identifiable basis upon which decision making is based).

This enables links and feedbacks between the financing and control of different projects and the change in the community determined by and affecting these decisions. This gives direct expression of the civic processes which act in conjunction with community needs and policy choices, giving an idea of the effects of change and the budgetary significance of alternatives and allowing for long term planning under specified objectives.

Comments

This model was originally conceived as an operational tool by which public officials can evaluate alternative policies on the basis of readily accessible information on the structure of the city. It is developed potentially for use in Metro Toronto.

So far only that part of the model related to the financial and administrative aspects has been developed, and it is possible that little more will be done due to a lack of financing.

It is difficult to evaluate the model due to a complete lack of information on the detailed structure and relationships. However it is notable in being one of very few attempts to explicitly include the public decision making process. There is one drawback to the particular approach used here; it is based entirely on the current administrative structure, with little thought for the possibility of developing alternative departmental structures which might lead to more efficient management of the urban area.

* * *

Putman, S.H. 1967. Intraurban Industrial Model Design and Implementation. Papers and Proceedings of the Regional Science Association 19: 199-214.

A model to predict growth in industrial plants and allocation of this growth (or decline).

Model

(a) Allocation of Employment to Sites

$$\Delta E^k_{i(t)} = \Delta E_{T(t)} \cdot (E^k_{i(t-1)} / \sum_{i=1}^{N} E^k_{i(t-1)}) \qquad i=1, . ., N = \text{No of sites.}$$

$\Delta E^k_{i(t)}$ = Change in employment in industry k, area i, from time (t-1) to t.

$\Delta E^k_{T(t)}$ = Change in employment in industry for the entire area, from time (t-1) to t.

$E^k_{i(t-1)}$ = Employment in industry k, area i, at time (t-1).

It was considered that no location effect was required because of the smallness of the area.

(b) Location of Declining Industry

If a decline was predicted of a percentage greater than that which an industry was considered able to accommodate without closing a plant; plants were chosen to be closed by a Monte Carlo technique.

(c) Location of Growing Industry

Sites were given attributes of assessed value, available land, structural density, and industrial clustering. Then the coefficient of each attribute is calculated as a weighted sum, over all areas, of the product of the size of the attribute and the percentage of a given industry present on each site. Then given these values, a new plant of each industry type would locate on the most favourable site; or if there were a tie for the maximum value a site would be chosen by Monte Carlo techniques.

(d) Generation of New Plants

If there were sufficient growth in an industry, the size of new plants was determined by Monte Carlo selection from the present size distribution.

Evaluation

The only part of the model evaluated was the location of new plants, and this was found to correspond well to observed location.

Comments

There is no inherent dynamic character in the model; the expansion of employment for each industry is determined exogenously and the model merely allocates it. However it was designed as only one of three linked models of the urban system, the other being an inter-industry input-output model, and J. Crecine's T.O.M.M.

* * *

Operations Research. 1972. Special issue entitled Urban Problems 20:3.

Includes bibliographies on work carried out by the RAND Corporation; and the Urban Institute, Washington, D.C.

* * *

Regional Planning Association of New York. 1962. Spread City: Bulletin 100. New York

Describes a simulation of intra-regional population change.

* * *

Rose, H.M. 1970. The Development of an Urban Subsystem: The Case of the Negro Ghetto. Annals, Association of American Geographers 60: 1-17.

A strict segregation-based model of the spread of the Ghetto in Milwakee.

Components of the Model

(a) Demographic component:

Growth of population is determined by births and death rates, and immigration for both negroes and whites.

(b) Producer:

Houses are assumed to be made available by white abandonment after a certain number of negroes move in.

(c) Consumer:

Households are formed by immigration and by employing marriage rates to the population. The directions in which they seek for housing depend on housing values and on distance (it was found that negroes usually moved less than ten blocks to a new house).

Evaluation of the Model

Two aspects were looked at: the change in certain blocks over time, and the overall spread after a certain time. The model tended to over-estimate the spread in areas of poorer housing, suggesting that migration in the model was too free; and tended to under-estimate total numbers of entrants in better areas, suggesting that the modelled propensity to move into more expensive housing was too low.

Levels of Aggregation

Spatial: Individual blocks are modelled.

Temporal: Each generation of the model represents one year.

Social: All negroes are assumed to act the same.

Comments

The model ignores the actions of white households in competing with negro households for housing before the negroes start to predominate in a block.

* * *

San Francisco Department of Planning. 1968. Status of the San Francisco Simulation Model

A model to simulate the housing market in San Francisco.

Components of Model

(1) Variables; households, housing units are finely divided, perhaps more so than anywhere else.

(2) Demand; based on projected household numbers and type.

(3) Supply; based on present housing, less those demolished.

(4) Pressure on housing is calculated leading to rent alternatives.

(5) The developments most profitable to the private investor are calculated on the bases of development costs and potential revenue, and these are developed.

(6) Updating of housing inventory and a further cycle of development if necessary is carried out.

Aggregation Levels

Spatial: location is not considered in the model.

Temporal: two generations are approximately equivalent to two years.

Social: a fine division of households is carried out, but all developers are assumed to act the same.

Exogenous Variables

Information on present households and population growth is necessary. Also needed is data on housing preferences for households (based on their present distribution between housing types), information on housing decay and rental levels, and on costs of land development.

Testing Results

Based on empirical data from 1960-1966 given the public policies at that time. For predicting numbers of houses of each type, only 1/4 of the results were within 10% of the observed values. It was concluded that the model should be scrapped as it would require a considerable cost to make it workable.

This may be due partially to the lack of consideration of the locatinal effects on development, and on the choice of sites to look at by households considering moving.

Demand for housing as reflected by growth in households is not related in any way to the supply that is available, theoretical content in general is lacking from the model apart from a simple economic land market formulation.

* * *

Schlager, K.J. 1964. Simulation Models in Urban and Regional Planning
The Land Use Simulation Model

A model to simulate the development of region.

(1) Residential

(a) Household decision:

i) To move; based on a turnover rate for each household time either deterministic or stochastic.
ii) Selection of new house type; depends on existence of housing and compatibility with this housing.
iii) Selection of location; depends on preference of each type for access to various facilities in the city.

(b) Developers decision:

i) How much land to develop, based on projected lot sales (a function of demand) and present vacancies in the market. -The date of housing being available is tagged by the building time from the date of land purchase.
ii) Where to buy land; linear programming to select sites so that total development cost is minimized.

(2) Industrial Sector

Given a set of sites which satisfy the demands of a particular industry type, sites are chosen to minimize total development costs.

(3) Service Sector

Allocation based on industry and population requirements.

(4) Special Sector

Public policies are added as specific inputs.

(5) Agricultural Sector

Seen initially as a residual land use from which everything is converted.

Aggregation Levels

Spatial: by local transportation zones of size 1/4 - 36 sq. meters.

Temporal: unknown.

Social: Household groups quite finely divided by age, income, etc. and firms by the Standard Industrial Classification.

Exogenous Variables

Household behaviour: from special survey, and similarly for industrial land requirements. Also needed are transport costs between all areas, data on soils and on building costs, on the quantity of development based on a service industries regression of demand over previous time periods which is extrapolated, and on size of population served by service industries.

Ability to Forecast

Based on altering public policies and industrial growth and examining results in a qualitative way.

Testing Results

To be based on just one town in the region but not yet carried out.

Comments

The model needs a large amount of special data which is usually only obtained once, giving an idea of how it changes over time; rather than trying to define such variables endogenously within the model. The relations are generally deterministic, which gives no idea of the variation from the deterministic results of running the model that might occur due to chance and unknown factors.

See Also

Schlager (1965).

* * *

Sears, D.W. 1971. Computer Simulation of a Regional Housing System, in Papers on the Application of Computers to the Problems of Urban Society, 6th Symposium. New York: Association of Computing Machinery.

A description of the New York State Regional Housing model of allocation of households to housing over time.

Variables

Households are split by age of the head of the household and by income into 30 types. Housing is split by value, structure and tenure into 18 types.

A matrix of preferences of households for housing types is constructed based on their present distribution.

Model

For each time period, the movement of households and housing between categories is determined by transition probabilities. Also the number of new households is exogenously determined based on probabilities of being in certain income groups. Then households are allocated to housing as best as possible resulting in an index of dissatisfaction and of excess demand, due to mismatching of the two categories. Supply of housing is then generated to meet demand. The amount of public housing built is given exogenously. On top of this, private housing is built such that all households are given a home. The type of housing built depends partly on the types desired most and partly on the present mix in the market. The final result is a market equilibrium achieved for five year periods in the model.

Comments

Location as a factor in preferences for housing by households is not included, and the categories used in general give a relatively coarse division of the variables. The validity of these features is debatable, The San Francisco model for instance used much finer divisions but achieved a poor fit to reality despite this. Another feature that is questionable is that an equilibrium situation is assumed; and possible changes in rental values due to excess supply and demand are not considered.

The model has as yet not been validated fully, although some tests that have been carried out so far are described as "encouraging."

The model presents possibilities of simulating alternative housing policies which could be validated by the size of the index of dissatisfaction (providing the assumptions under which this has been derived are valid).

* * *

Silvers, A.L. and Sloan, A.K. 1965. A Model Framework for Comprehensive Planning in New York City. Journal of the American Institute of Planners 31: 246-251.

Models

(a) Programming Policy for Low Cost Housing

Choice of that programme which maximizes the product of the number of houses and their value characteristics in a linear programming format subject to budgeting and space constraints.

(b) An Investor Behaviour Model

A simulation of the behaviour of landlords in the land market is proposed, based on postulated curves of rent, taxes, maintenance costs and financing costs graphed against housing conditions. Here the entrepreneur acts to maximize profits, where revenue and cost are themselves seen as functions of environmental conditions and government constraints as well as housing conditions.

(c) A Residential Location Model

Here the attraction of an area for an income group is the sum of neighbourhood characteristics weighted by the group's perception of the desirability of such characteristics. Then the probability of moving to a particular site with certain characteristics is equal to the fraction of the total market in these types of sites that this site takes up.

Then it is proposed to model changes in these characteristics as a result of public policy, leading to a resultant change in the attractiveness and hence demand for these sites.

The overall aim is to combine these models into a framework where the effects of certain policies in economic development on the demand and supply of housing types to particular income groups can be simulated. This in turn may lead to suggested programmes to help carry out social policies such as slum reduction.

Comments

The models are in a conceptual stage only, and the possibility of calibration based on present data availability seems a little remote. Until a stage of evaluation is reached it is difficult to comment on such a model.

*   *   *

Stéger, W.A. 1965. The Pittsburgh Urban Renewal Simulation Model. Journal of the American Institute of Planners 32:2;144-150.

Components of Model

(1) Employment Opportunities

These are directly or indirectly responsible for all development. Employment projections are expanded to labour force participation (via an input-output interindustry model) as a result of multiplier effects of projected sales which in turn leads to journey to work patterns and population patterns, which are compared to separate area by area direct population projections. A compromise value, but dominated by employment projections, is used.

(2) Allocation of "Basic" Industry within the City

Based on site attractiveness measured by industry already present, land use policies, access, etc. Industrial requirements are relaxed and policies are changed until either a site is found or the industry leaves.

(3) Allocation of Residentially Oriented Industries

Given the employment and journey to work areas for "basic" industries a pattern of population results which determines the location of these service industries. This induces further population employed by these industries in turn, leading to a few more service industries - a process

repeating until the model is approximately stabilized. The allocation is constrained by population already there and by zoning policies.

(4) Reconciliation

The modelled results are compared with independent projections of the variables.

(5) Parameters Determining Model Structure

The determinants of features such as household mobility, housing types demanded by households, locational behaviour of industry, and others are calculated by factor analysis from an original number of up to 600 variables.

Aggregation

Spatial: by census tract.

Temporal: unstated.

Social: a large number of linear independent variables are considered.

Exogenous Variables

Data is required on structure of the households, on coefficients for the input output model, on projected growth in sales, on public policies, on tributary populations for service industries, and on locational behaviour and minimum size of all industries.

Output

Results by census tract give: population, employment by S.I.C., land use social indices, inputs of public projects, housing, surplus and deficit markets for firms, income, new firms, blight indices (calculated from measures of deterioration).

Also for the city as a whole, information is obtained on revenues, expenditures, fiscal alternatives, and project priorities. Normative planning is possible but no explicit criteria to evaluate alternatives are given.

Testing Results

Not carried out.

Comments

The description is too general to allow much criticism, but possible shortcomings are that all relations of causal variables are linear and all coefficients and relations are assumed constant throughout projections. The limitations of these assumptions, for instance for long term projections

of input output models, are well known. Also an extremely large amount of data seems to be used to calibrate the model.

See Also

Steger (1964).

*　*　*

Steinitz, C. and Rogers, P. 1970. A Systems Analysis Model of Urbanization and Change

A model developed to simulate the growth of the Boston hinterland region. The model is essentially a gaming approach. Models of allocation of industry, housing, commercial centres and recreation areas are developed; using indices of site attractiveness calculated by regression equations and supported by subjective judgement. Any conflicts arising are solved by discussion between the different groups. In addition models of evaluation of the effects of activity allocation based on political, fiscal, pollution and site appearance (visual) factors are developed and act as constraints under which the allocation proceeds. Thus if conflicts come up between say industrial location and the desires of the people to take industry in their town (expressed as political pressure) then these would have to be resolved by discussion and bargaining. As more generations were carried out less conflict appeared in the model, as actors steered away from conflict areas using their experience of what the outcome would be.

*　*　*

Swanson, C.V. and Waldmann, R.J. 1970. A Simulation Model of Economic Growth Dynamics. Journal of the American Institute of Planners 36: 316-322.

Model

A model to simulate the changes in labour force, employment and population in a region over time. Based on the precept that regional economic growth is a non-linear dynamic interaction process between the job and population sectors. The model is built on the concepts of dynamic non-linear feedback systems developed by Forrester.

(1) Population is divided by age and occupation. Formulae for movement within this classification due to aging, migration, and occupation and participation rate changes are set up, for the various occupation groups, e.g. inflow into an occupation cell = + (overall unemployment rate - average unemployment rate for this occupation).

(2) The job sector is split into three parts which are in turn split into smaller groups.

a) Manufacturing sector

Increase in jobs here is a product of factors representing: labour availability, national growth of the industry, local jobs in the industry and the differential advantage of this region for the industry. Parameters for these are taken from graphs as for the Forrester model.

b) Domestic sector

This is just an exogenous function of time. Local fluctuations are assumed to have no effect on this industry.

(3) Linking the sectors is carried out by postulating graphical relations between availability of labour and the fraction of labour employed.

Validation

This was carried out for Kent State and results we found to be comparable in many respects (though over the same period for which the model was calibrated!). The model tended to show a damping in fluctuations because the future values were based on projections of exogenous variables which themselves are smoothed by their nature.

Comments

Equations and relations are stated without giving any theoretical principles as rationale. Spatial effects are ignored as are the government sector, the land use and transportation sectors, and education. The population is split by age and occupation only. The type of concept is useful but needs to be formulated and related to theory before any policy implications of its results can be accepted with confidence.

* * *

Taaffe, E.J., Garner, B. and Yeates, M. 1963. The Peripheral Journey to Work: A Geographical Consideration. Evanston: Northwestern University Press.

A model is built to simulate distances of the journey to work, traffic densities on routes based on empirical data for Chicago.

* * *

Tobler, W.R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46: 234-240.

A model to provide prediction of population growth.

Model

Assumptions:

Growth is influenced by nearby cells and is estimated as a linear function of this. The growth and migration of population is assumed constant for all cells.

Initial Model:

$$P_{ij}^{(1930+\Delta t)} = \sum_{P=-2}^{+2} \sum_{q=-2}^{+2} W_{pq} P_{i+p,\ j+q}^{1930}$$

$P_{ij}^{(1930)}$ = Population in cell (i,j) (the map is expressed in matrix form)

W_{pq} = The average spread function of a population unit (the sum of population growth and migration).

Where

$W_{pq} = A_{pq} + B_{pq}\,\Delta t$ where A_{pq} and B_{pq} are regression coefficients calculated by regressing the population at time (t = T) against the population at (t).

Giving an intuitive rather than formal insight into the dynamics of urban growth.

<u>Evaluation</u>

Subjectively:

a) Underestimates decline in the C.B.D.

b) Tends to lead to excessive smoothing of maps, due to only specifying a neighbourhood influence over two cells in each direction.

Suggests that functions for W_{pq} that are locationally invariant are not a good thing; modification with distance from the C.B.D. would be better.

<u>Comments</u>

An empirical approach; forecasting based on constancy of rates of change of conditions. There is no theoretical structure by which to incorporate of true dynamics.

* * *

Walsh, B.F. and Grava, S. 1969. A Dynamic Land Use Allocation Model. Papers on the Application of Computer to the Problems of Urban Society, 4th Symposium, Association of Computing Machinery, New York.

A model which simply allocates land uses to land plots depending on certain desires.

Exogenous Information

(a) Presence or absence of characteristics (or of a certain level of a characteristic) on each lot, expressed as binary numbers.

(b) Preference function for each land use specifying those characteristics which are necessary or inadmissable, and those which are more or less desirable.

(c) Land use and changes in land use during the study period for each lot.

(d) The growth in demand for each land use.

(e) The length of time a land use can exist before it declines and is replaced.

Model

For each generation, each land use looks for feasible sites such that the convenience of the site chosen (expressed as the goodness of fit of preferences to actual characteristics present) is maximized. Then given the number of units to be allocated they are put on the "best" sites.

Thus the order in which different land uses get a chance to seek for sites must be specified, as the first one obviously has the best choice (the real world situation is not dissimilar).

Comments

Levels of aggregation in space, time and in the number of land uses postulated are variable. The model is as yet untested.

If land uses do not find sufficient sites, the unsatisfied demand is noted; but no attempt is made to represent a change in either land use due to market pressure, or in the preference functions of land uses due to not finding a location.

Continuous variables, such as locational preferences based on distance, are difficult to convert to the binary format used. A lot of data on preferences, etc. is needed for what is essentially a very simplistic view of the allocation process; in order to match the real world situation (because no theoretical rationale is postulated).

* * *

Wärneryd, O. 1968. Interdependence in Urban Systems

The Basis of the Approach

The urban system is here seen as a time - space system based upon the organization theory of systems. Under this the urban system is seen as being based on a series of public and private organizations, each possessing a

hierarchical structure of plants with several local level plants distributed through the regions of a country each under the rule of and hence interacting with a smaller number of "intermediate level" plants which in turn are connected to a single national plant. It is postulated that for different organizations plants at a certain level tend to agglomerate in certain regions for convenience of interaction with the other organizations, and in particular most head offices of these organizations tend to locate in the same area. The urban system and its interactions are based on these agglomerations and the relations up and down the hierarchies of the organizations.

The resultant interactions between the urban areas, which may be measured in many ways, are seen as specifying the interdependences between these areas. This interdependence means that when an impulse or innovation is applied at any point there is a resultant spread effect to other areas with which it is connected; the size of the effect depending on the relative position of the other areas in the hierarchy. (It is assumed that different organizations have their structure set up in a sufficiently similar way that the cities in which these organizations have plants can be assigned fairly unambiguously to a local, intermediate or national level of the hierarchy. Some empirical work has shown for Sweden that this may be reasonably valid.) Based on this idea a temporal dimension can be given to a static central place type system based on innovation (interpreted in the widest sense as any change in policy by an organization), interaction and interdependence. Thus the change in policy by public and private organizations is seen as the major determinant of urban change. It is postulated that the policy folllowed by any plant is first of all expansion and regionalization as the firm grows and takes over others, and then a contraction and increasing concentration of operations as automation takes over; but different organizations are at different stages in this cycle at any one time.

Model

(a) Location of Innovations

The probability of an area accepting an innovation is proportional to its population. Innovation is interpreted solely as the opening or closure of plants.

(b) The Interdependence of Areas

The probability of one area affecting another is proportional to the relative interaction measured between those areas (based on rail traffic, parcels, telephone calls).

(c) The Spread Effect

A three by three matrix of effects of innovations at one level of the hierarchy on itself and each of the other levels is drawn up based on subjective judgement due to lack of data. Different matrices are constructed for positive and negative effects - the latter being assumed to only flow down the hierarchy.

(d) The Hierarchy of Towns

This is based on the dominant interaction flow from a town which is assumed to be an indication of which place this particular town is immediately subordinate to in the hierarchy.

(e) Rules

Three random numbers are drawn

i) The location of the innovation is chosen.

ii) The second region affected by this is picked and the result evaluated from the effect matrix.

iii) A secondary effect as a result of the stimulus to the region chosen in (ii) is calculated by choosing a third based on its interdependence with the second one.

iv) This is subject to two rules

a) If the spread effect is chosen to be from one intermediate level city to another, the size of the effect is as if the interaction were between an intermediate and a local level city.

b) If the spread effect is between two local level cities, there is no stimulus travels to the second local level city, but stage (iii) is still evaluated relative to this second city.

These rules are a result of organization theory which specifies that there is only interaction if one region is at least partly under the jurisdiction of the other. Thus one intermediate level city can only interact with the local element in another intermediate city, and two local level cities cannot interact.

Empirical Validation

The model was run for sub areas of Sweden, first allowing only positive stimulus and then allowing a negative element to occur as well (it was assumed that every second time a local city picks up an innovation and every tenth time an intermediate city does, this will be a negative stimulus; an arbitrary rule). Then the model was run for Sweden as a whole. Each time the model was run to allow a number of innovations approximately equal to what might occur over a period of several years. But each such run was only made once. The results were compared with net migration data for the regions, since a unit stimulus is assumed to give rise to a given amount of immigration, and the fit was found to be poor. It was discovered that altering the hierarchical structure or the values in the spread effect matrix (particularly the very large elements on the main diagonal) affected the results.

Comments

There are two major points to be made.

(a) This particular theory of urban change is difficult to test as the concepts within it are not related directly to easily available data. This is

because there are little data on the policies undergone by all the organizations even in Sweden. Thus stages (a), (c) and (d) in the model construction and also the evaluation against the real world have to be extreme surrogates for the true data. This makes it very difficult to know when the model is working. The assumptions involved here are at least as severe as in the model itself.

(b) There are several shortcomings in the transformation of theory into model. It is assumed that a region with an innovation affects only one other region directly with one other being affected in the second round. This would be allright if such an effect were an indivisible unit, but there is no reason why an intermediate centre might not stimulate several local centres. Also there is no reason why more than one round of indirect effects should not occur. Techniques such as the input output table and the Lowry model show how to deal with this. In addition the second round effect is assumed to be as big as the first round effect which is incompatible with the idea of there being no third round effects - one must at least assume a fast drop-off in effect. Also only one run of the model is made and the fit of this run is very dependent on the number chosen.

The temporal element is still somewhat arbitrary as there is no relation to real time. Finally the rules prohibiting interaction at one level of the hierarchy may be incompatible with real life where contiguity as well as the hierarchical effect plays a role in regional interdependence. However it is one of the few attempts to model intercity effects and also employs a slightly different philosophy towards urban change as its basis and for these reasons this study repays examination.

* * *

Wolfe, H.B. 1967. Models for Condition Aging of Residential Structures. Journal of the American Institute of Planners 33: 192-196.

A first order Markov model is used to predict transition of houses from stages of better to those of poorer condition. "Considering the nature of the data . . . (for) . . . buildings constructed over 50 years the fit of the theoretical curves to the data is quite good."

The model could be generalized to look at results under alternate public programmes, but the data to estimate transition probabilities would be difficult to obtain.

DYNAMIC SIMULATION MODELS OF URBAN GROWTH: AN OVERVIEW

ERIC SHEPPARD

For the purpose of this review, papers on the application of simulation techniques to the modelling of urban change were searched for in recognized journals and other widely available publications. Papers produced by individual research bodies that have only limited circulation were in general not looked at although many of these have been included in the bibliography.

The review is divided into two parts: first, the structure of simulation models and the uses and limitations of this structure for urban modelling is discussed; and then a classification of those papers looked at is attempted. Before proceeding with these, however it will be useful to define the type of model reviewed here. Kilbridge, et al. (1969) define simulation models as follows:

> . . . . simulation refers to the way in which a model is used rather than to the structure of the model itself. However, it is sometimes possible to distinguish simulation models from analytic models on the basis of mathematical structure. Analytic models contain precise mathematical statements which can be solved - at least theoretically - by standard mathematical operations. But simulation models usually include nonmathematical statements about relationships between elements which, though sufficiently clear for conversion to a computer program, have no mathematical 'solution'.

This is a reasonably accurate characterization of what is referred to in the urban literature as simulation. However, in practice both models which are purely analytic, but require empirical relations from data for the computer to solve them, and models which contain some purely subjective elements dependent on decisions by the operator are also included. Only one or two examples of the latter "gaming simulation" models are included

here, as the majority of such models are at present used for primarily educational purposes, rather than as serious attempts to model the urban system.

1. The Structure of Simulation Models

1.1. The Concept of Generations

Most simulation models centre around the idea of successive generations, in order to incorporate dynamic structure. This concept means that the model is solved anew for each successive time period, where the solution for the next time period depends on the structure produced during the previous time period. There are two major properties of interest here. The first is that the length of time represented by each time period is arbitrary, and in general it is difficult to state whether the model should be solved once for each year, say, or once for each month. So although the model is flexible in this sense, it is often difficult to interpret results and to compare them to the processes operating in real life, unless one has some idea of the rate of change. The second property is that generations make the model incremental in nature. This means that the time-path of the model can be examined by looking at the results of successive generations, and also that, if necessary, changes in conditions can be introduced at any point in time. It is this property that makes simulation models most attractive as a way of modelling the dynamics of a process.

Within each generation, the model is solved either simultaneously or sequentially. Both methods are valid, depending on the process being modelled and the length of time of the generation. Obviously when this time is very short sequential solution would be inappropriate. To specify a sequential solution there must be a recognized causal order by which each variable

affects the next one.

1.2. The Mathematical Structure of the Model

Usually the model is constructed as a computer programme which is flexible in the sense that the values of parameters can be altered with ease. As all the work is done by the computer it is easy to test the sensitivity of the model to changes in individual, and groups of parameters which in turn gives an idea of how to alter the model to achieve certain results.

Solutions can be either deterministic or probabilistic. Probabilistic models have not often been used in urban situations, but there is great potential here. This is because even when there is no analytic solution to the stochastic process proposed, it is always possible to run the model many times and obtain results for the average run and the variance in the runs about this average.

It is relatively easy to include subjective elements in the model - steps which depend partly on decisions of the operator - because the incremental structure of the model allows intervention at any point, and also because the explicit structure of the model means that the variables in the model upon which such decisions depend are easy to pick out.

1.3. Testing and Evaluation of the Model

1.3.1. Testing:

The explicit structure of the model means that the data needs for calibration and testing are known, and these just need to be fed into the computer programme to give post-dictive and pre-dictive comparisons with the real world. Objective testing of the goodness-of-fit however is difficult. Testing of simple frequencies of modelled and observed data to

give an idea of the fidelity of the model is relatively well developed, but the testing of spatial patterns is still in its infancy. Statistics developed can test whether the deviations of the modelled pattern from the observed one are random or not; but even when deviations are random they can be so large as to be of little use in prediction of future patterns, and we have no way yet of saying when such deviations become significantly large.

1.3.2. Evaluation:

In order to evaluate alternate runs of a model, such that an idea can be obtained of what is an "optimal" run, there are needs for the specification of some evaluation function. The problem of specifying such a function in terms of true cost to society is a difficult and as yet unsolved one which remains vital for the eventual use of these models in planning the future structure of cities. There is a lot of work to be done here for all types of urban planning models, not only simulations. However once such a function is produced in some form, it is easy to feed it into the model and have the computer calculate its value over a set of alternative runs.

2. A Classification of Simulation Models

In this classification the models have been split into two groups: the planners' models, which are those developed to solve specific problems in certain cities and which have been used or are to be used as a direct aid to planning; and the academics' models which have been developed in research institutions and tend to be of more general applicability. A further subdivision is made into applied models, which have been calibrated with or tested with available empirical data, and potential models which have not reached the stage of calibration and testing. Within these four main groups the models are split up according to that part of the urban system which they model, in order to give an idea of which subjects have received more or less attention.

2.1. Applied Planners' Models

2.1.1. Industrial Location:

The Bay Area Simulation Study (Centre for Real Estate and Urban Economics, 1968; Nathanson, 1970; Walter, 1968) was developed to study the impact of new industrial plants on residential and commercial location, using a version of the Lowry model. From here was developed the Projective Land Use Model (Goldner, 1968, 1972) a refined version where vacant land was no longer a residual in the allocation process, and where the gravity function of interaction was modified.

2.1.2. Urban Renewal:

This was built to test the effect of alternative building policies on housing availability.

The San Francisco Model (Little, 1966; Robinson, et al., 1965, San Francisco Department of City Planning, 1968; Wolfe, 1967) used housing aging (by a probabilistic process) and household change with time to predict future demand and fulfilment of this by private housing under alternate public policies. It is the most highly disaggregated model of its type, but is aspatial, and the San Francisco Department of City Planning decided in the end that the model was too inaccurate to use (San Francisco Department of City Planning, 1968).

2.1.3. Land Use Allocation Models

The Pittsburgh Model (Steger, 1965) worked from employment projections by subarea to give future population and location of basic and population oriented industry. The model was solved as a simultaneous solution with parameters based on factor analysis of Pittsburgh data.

There are a series of other models; for Boston (Hill, 1965) - based

on multiple regression and simple extrapolation for forecasting; for Philadelphia and for Connecticut (Lamb, 1962) - econometric type models; for New York (Regional Planning Association of New York, 1962) for Wisconsin (Schlager, 1964, 1965) - a model based on data for household selection of new housing and for industrial land requirements, using regression and linear programming; and the Penn-Jersey Transportation model (Irwin, 1966) - an attempt to look at the relations between land use and transportation demands.

All these models depend on simultaneous equation type of solution for each generation. Generations are of arbitrary length with no rationale for choosing this length. The evaluation and testing of the models is largely unstated and this tends to be true of all the planners' models, which makes it difficult to discuss their use in practice.

The only model built outside the United States (McLoughlin, 1969) is difficult to classify with the others as it employs a different approach. Many of the decisions of where to locate activities, given the present state of variables in the system, are made subjectively by the operators based on their discussion with planners in the region in question. This is an attempt to copy the actual decision process rather than to try and represent it in equation form; an interesting alternative approach although its validity remains to be seen.

Many of the above models show little in the way of true dynamic character; the relations are kept constant and each generation tends to be a form of extrapolation based on these relations and on the existing structure. There seems to be no attempt to incorporate changing relationships such as say, the business cycle. In addition all relations tend to be linear.

Housing is covered in all models, industry in many, but transportation is only one in any real endogenous form. It tends to be added in as a sort

of after-thought. Also the models tend to be based on high data needs often requiring, or taking advantage of special local surveys. The high data needs also is often complemented by relatively low theoretical content, taking the form of single relations and regression equations. Several of the models ignore the spatial aspect as well. These things may account for the poor behaviour of the San Francisco renewal model, despite its great detail. (This is the only model for which there exists a careful evaluation.)

2.2. Potential Planners' Models

As one might expect, this is a very small group.

2.2.1. Housing

Silvers and Sloan (1965) propose a model of residential location based on potential household behaviour with respect to housing characteristics, looking for policies that will maximize the social value of a housing programme. But the model is very difficult to operationalize and calibrate because of lack of behavioural theory.

Sears (1971) discussed what is in effect a simplified version of the San Francisco renewal model, but there is no reason why it should work any better. The model is proposed for use in New York State.

2.2.2. Transportation

The Northeast Corridor Transportation Project (1970) proposes a simulation model of the impact of highways.

As far as these potential models show the future orientation of planners, there seems to be little change from the 1960's. None of the planners' models have attempted to model the inter-urban relations specifically, although some are at the subregional scale looking at a major city and its hinterland.

2.3. Applied Academic Models

2.3.1. Housing

This represents the only substantial group of probabilistic models. These have stemmed from two sources:

The University of North Carolina model (Chapin, 1965; Chapin and Weiss, 1962, 1965; Donnelly et al, 1964) was developed as an allocation model where the probability of development of a site depends on attractiveness, which is calculated by "a best fit" regression on site characteristics. Then random numbers are drawn to allocate units based on these probabilities. Another model coming from the same source is that of Massie (1969) which uses the same type of allocation principle, but calculates attractiveness based on the effect of landowner, developer, and site characteristics on the probability of site development. In other words it takes a more behavioural approach. However both of these models consider supply of housing primarily rather than demand. Also the probabilistic element is not really built into the model as part of the process of determining site attractiveness.

Morrill is the only geographer to have worked explicitly with simulation until very recently. His model of the expansion of the urban fringe (Morrill, 1965a) uses site characteristics alone to construct probabilities of development. Functions such as the probability of development with distance from arterial roads, are calibrated for Seattle and the model is then run using data for the same area using these functions. Not surprisingly, the results give reasonable patterns.

The other work by Morrill (1965c) at this scale involves the expansion of the negro ghetto; based on migration distances of negroes relocating and on decreasing resistance of blocks to negroes as the number of negroes in that block increases. Rose (1970, 1972) has developed a similar model which in

addition incorporates the effect of housing values on negro in migration. Both models are based on the principle that white exodus gives rise to vacancies and that there is no competition between negroes and whites for the same housing. This seems a very simple assumption which might be worthy of modification.

The only Canadian model in this entire classification is that of Larouche (1965) for Montreal, which attempts to simulate residential development.

2.3.2. Transportation

Ford and Jago (1968) provide one of the very few attempts to consider evaluation of alternative policies; they consider the costs and benefits of alternate transportation mixes in some U.S. cities, projected forward to 1990. Taaffe et al (1963) consider the journey to work in Chicago. Colenult (1969) looks at the effect of road traffic on billboard construction.

2.3.3. Urban Growth at the City and Regional Scale

Hester (1970) models the effects of decentralization, segregation and technological change in large U.S. cities. Steinitz and Rogers (1970) take an alternative approach of a more subjective nature, similar to that used by McLoughlin (1969). Here, location of activities is based on decisions by operators representing each type of land use; and modified by discussion between these, whenever there is a conflict of interest between groups or between a group and local government. This is a very simple model with little attempt to describe relationships in a precise way.

Morrill (1962, 1963, 1965b) builds a probabilistic model where urban growth depends on transportation, industrial location, the laws of central

place theory, and migration. It is a comprehensive model conceptually, but falls down in practice as it assumes simple relations, and as there is no attempt to calibrate it with real world data when it is tested (for an area in southern Sweden). It also only tends to hold for areas where central place laws predominate over, say, location of resources as a cause of city location.

Swanson and Waldman (1970) construct a model of regional economic growth based on the complex systems approach of Forrester (1969). The model uses population, manufacturing, primary and domestic activities as the basic variables with graphic relations postulated (although not empirically validated) between them. The model uses national and local trends to give changes in industrial sectors which in turn gives rise to labour force, employment and population projections. Subjective testing of the model against thereal world produces 'reasonable' results. However, there are no spatial elements in the model, and relations are not calibrated against real world data. This type of approach may be useful in modelling the response of cities to the national economy and the economies of other regions. Tobler (1970a) produces a simple extrapolation model of population growth by sub-area in Detroit, based on an auto regressive scheme whereby population growth in each area is affected by present values in nearby areas, as well as in the area in question.

These academic models still show surprisingly little theoretic content, with the exceptions of Swanson and Waldman, and Morrill (1965b). And the two latter studies tend to err on the other side. Objective testing is not given careful consideration (with the exception of Colenutt), despite its vital role in the process of testing these models. Given that this is a group of academics presumably well versed in the principles of modelling the real world, this lack of attention to testing is disappointing.

2.4. Potential Academic Models

2.4.1. Housing

There are two probabilistic models that come into this classification. Maln, et al. (1966) built a model where probability of apartment development depends on site characteristics as they affect construction costs. Drewett (1969) develops a semi-Markov model of land use change where probabilities of change are modified by the duration of each plot in its present state. This depends on extrapolation based on constancy of rates of change, in order to predict into the future.

There are a whole series of deterministic models of housing allocation and choice (deLeeve, 1972; Graybeal, 1966; Harris, 1969; Pilgrim, 1969; Walsh and Grava, 1969;Wheaton and Harris, 1971). Walsh and Grava represents the simple allocation approach to minimize dissatisfaction of households (with supply and demand given). Pilgrim considers the demand side only, looking at the effect of household behaviour with respect to property characteristics on the choice of housing. Wheaton and Harris consider both supply and demand in a model that is based on the economics of the household, and seeks to maximize household satisfaction. Graybeal also considers both supply and demand.

2.4.2. Commercial Location

Cowan, et al. (1967) model the supply of offices based on the numbers of new and converted buildings appearing on the market and the time taken to build - all seen as probabilistic processes within the model rather than just in the allocation procedure. However the model is aspatial; ignoring the effect of site characteristics on location.

Harris (1964) develops an equilibrium model of the location of

retail trade outlets. Putman (1967) looks at industrial location picking out factors that determine the location of plants that are closing down or opening up. The demand side of the model's expressed as the change in employment; which is a function of the total change in employment in the area. It gives forecasts for each area. This is part of a model developed by RAND to study the relations between land use and transportation. (See also Kain, 1965.)

.4.3. Transportation

Creighton Hamburg Inc. (1969) looks at the relation between transportation and land development, as does Garrison (1966). Nystuen (1967) builds a simulation model of urban travel behaviour.

.4.4. Models of the City as a Whole

Models of city growth and structural change have been approaches in a variety of different ways. Crecine (1966, 1967, 1968, 1969) has developed a dynamic model based on the Lowry formulation. It differs from Lowry in that each generation builds on the structure of the one before rather than starting with an area of empty land and building a city on it all in one period. This is done by defining a certain percentage of households and firms as mobile in one time period and the rest as static. Then for each generation; given a growth in exogenous (non-population oriented) industry, this gives rise to location of population to work there, which leads to population oriented (endogenous) industry to serve the people. This in turn needs more people to work in endogenous industry which means more of this industry to serve these, and so on; interating until a balance is achieved. The locations are based on specified preferences for individuals. The later model (Crecine, 1968) also includes site amenities as well as

location as factors in evaluating preferences; and zoning and other constraints to introduce greater realism. The model has great data requirements and has not yet been used. Batty (1971) has produced a model that is similar in conception but which includes a specifically dynamic hypothesis; that the effect of exogenous industry on subsequent location of other activities is not all dealt with in one generation but also has a significant but decreasing influence in the next few generations as well. So there is no attempt to move to an equilibrium within each generation. However functions affecting population and endogenous industry location are simpler than in Crecine's model, being based on location and access (with access measured by the potential model) characteristics of land plots alone. In both these models there is no attempt at calibration, which could be quite difficult, and no attempt to predict the growth in exogenous industry.

Engle, et al. (1972) try an alternative of an econometric model. This is split into two parts. The first determines population, capital investment, employment, production, and income based on the levels of economic and demographic variables within and outside the city. The second attempts to allocate these activity levels to subareas within the city, as a market clearing process. However no attempt is made to state the form of functions in the model, or even the exact variables that will make up such functions in some cases. Also, the idea of reaching a market clearing equilibrium in each generation seems an over-simple description of the process.

Blumberg (1971) attempts to integrate several other models of parts of the urban system into a general model of the community, which will specify changes in all the socio-economic variables which are useful in evaluating the way of life in subareas of a city. The models put together are the

Penn-Jersey transportation model (Irwin, 1966), the Pittsburgh land use model (Steger, 1965) and the San Francisco housing model (Little, 1964), but no description is provided of how this is done or of how well it works. It is also hoped to include public policy endogenously as another linked model.

The most original approach to this problem is by Forrester (1969). Working from his own concept of complex, non-linear systems he defines a city of fixed small land area, in an environment of unlimited resources (migrants being the major such resource) where the environment affects the city but the city in no way influences values in the environment. The model connects a series of state variables (levels of housing, industry, population) by a set of rates or flows which depend on previous levels of the state variables. The model has all relations in a non-linear form and introduces the interesting concept of reading values of some functions directly off non-linear graphs, which could be empirically constructed. However the functions he uses are not related to urban theory or empirical data. The model then runs for 250 years after which equilibrium is achieved irrespective of initial conditions. The model makes no attempt to locate activities spatially within the city, as it is assumed that there is no spatial segregation of activities. Because the model took a radically new approach, independent of any previous urban theory, it has received much attention and criticism (Fleisher, 1971; Horton and Morris, 1970; Moody, 1970; Newling, 1970; Tobler, 1970b; IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-2, No. 2, 1972). Also there has been some work done with it; Garn (1971) has made some modification and criticism of the earlier model;

Graham (1972) has built suburbs into the model as a second region, but with no real change in the model; and Stonebreaker (1972) has shrunk Forrester's version a simpler model which behaves the same way. But much work must be done in making the model realistic, before its utility can be evaluated.

2.4.5. Inter-Urban Models

Lathrop and Hamburg (1965) have produced a model where the opportunity for growth at any point in the urban system depends on accessibility in travel time to the central place of the system. Thus the amount of growth is ordered according to increasing temporal distance. Also the model is such that as the probability of development of any given unit of opportunity falls, then overall development in the system becomes more dispersed. This simple model assumes that development depends entirely on access on the central place.

Kadanoff and Weinblatt (1972) are developing an inter-urban model based on the Forrester model. Each city is split into suburbs and city centre, and migration of firms and people between city centre and suburbs, and between different cities depends on population and relative attractiveness of areas. Also an explicit, though simple, demand function is built in to determine industrial growth dependent on total population in the nation, and location of the non-local employment sector is specified (as a function of city size and accessibility to the market). The ideas behind this approach are useful but its utility depends on the validity of the inter-action and allocation functions. Also there is still no attempt to link empirical data to Forrester's postulated functions.

The group of potential academic models is the largest one and also is the group showing the widest variety of different approaches, partly

because it is far easier to build an intuitively reasonable model than to test it. A result of this is that many of these models have huge data requirements in order to estimate parameters, which makes future use of these seem very remote. Few people seem to be working towards the compromise of reasonable data requirements with not too much abstract, unfounded theorizing. Although this is a difficult aim, as high theoretical content does not necessarily imply low data needs, it should at least be attempted if we are ever going to produce models of general applicability.

As far as coverage of the urban system goes, even conceptual models of government policy and response to urban problems have not yet been developed. Also transportation models have still to be integrated with other models, without assuming that development is caused primarily by transportation links. The other problem which is still getting too little attention is that of linking future growth and change to projections of the national and regional economy.

3. Conclusions

From the literature available, it is evident that simulation models of various kinds have received much attention as a possible method of predicting the growth and change in urban areas; and among planners in particular there was a phase in the 1960's of widespread popularity of these methods. It is worth, however, reviewing at this stage how much has truly been achieved.

As far as modelling the entire urban environment goes, there are several aspects which have received scant attention. The bulk of the models have dealt with land use and structural features in the city; and in particular much work has been done with housing and commercial activities in predicting their growth and location. Transportation has also received attention though

a widespread integration of this with the rest of the urban system has not really been achieved. However public facilities and their location and also government policy as regards development have only been included as exogenous inputs, rather than as a part of the system which is partly determined by the present state of the city.

Another area which has received very little attention is the prediction of the sociological variables describing the way of life and standard of living of the urban population (apart from the obvious variables of employment, occupation and population size).

Most of the work so far has been at the intra-urban level which has led to detailed models often with enormous data requirements. Work at the inter-urban level has different requirements with conceptually, needs for data that would be less disaggregated. However work at this level is still at an early state; and in particular little work has been done on the dynamics of the connections in an inter-urban system and its relations to the business cycle and other aspects of the national economy.

The final element missing from the majority of models is any form of evaluation to give an idea of the costs and benefits to society of alternative programmes.

There are a few other points about the structure of the models that come to mind. First, the models, with the exception of that by Batty, tend to be Markovian in the sense that only the structure in the previous period will determine the present state. One might expect a truly dynamic model to include the effects of earlier periods more explicitly. Also many of the relations in the models are of a constant form over time even though the spatial structures of the models is changing. In this sense the models tend to be a form of static dynamics. Probabilistic, as opposed to

deterministic, models have not been much used, despite the ease of incorporating probabilistic statements, and the large-scale acceptance of the notion of probabilistic rules affecting behaviour. Finally, there is very little evidence of experimentation with the models, such as an analysis of the sensitivity of the models to parameter changes. Such analyses are easy to carry out, and essential if an idea is needed of which planning strategies should be applied to give maximum impulse toward desired goals.

FOR REFERENCE ONLY
POUR RÉFÉRENCE SEULEMENT

7975